Epik 1.0

User Manual

Schrödinger Press

Revision A, April 2006

Contents

Document Conventions vii

Chapter 1: Introduction 1

1.1 **The Epik Manual** 1

1.2 **Epik Technology** 2

1.2.1 pK_a Prediction 2

1.2.2 Structure adjustment 2

1.3 **Installation and Licensing** 3

1.4 **Other Utilities** 3

1.5 **Citing Epik in Publications** 4

Chapter 2: Introduction to Maestro 5

2.1 **General Interface Behavior** 5

2.2 **Starting Maestro** 5

2.3 **The Maestro Main Window** 6

2.3.1 The Menu Bar 8

2.3.2 The Toolbar 9

2.3.3 Mouse Functions in the Workspace 12

2.3.4 Shortcut Key Combinations 13

2.4 **Maestro Projects** 13

2.4.1 The Project Table Toolbar 15

2.4.2 The Project Table Menus 16

2.4.3 Selecting Entries 17

2.4.4 Including Entries in the Workspace 17

2.4.5 Mouse Functions in the Project Table 18

2.4.6 Project Table Shortcut Keys 19

2.5 **Building a Structure** 20

2.5.1 Placing and Connecting Fragments 20

2.5.2 Adjusting Properties 22

2.5.3 The Build Panel Toolbar 22

2.6 Selecting Atoms 23
2.6.1 Toolbar Buttons 23
2.6.2 Picking Tools 24
2.6.3 The Atom Selection Dialog Box 25
2.7 Scripting in Maestro 25
2.7.1 Python Scripts 25
2.7.2 Command Scripts 26
2.7.3 Macros 27
2.8 Specifying a Maestro Working Directory 27
2.9 Undoing an Operation 28
2.10 Running and Monitoring Jobs 28
2.11 Getting Help 30
2.12 Ending a Maestro Session 30
Chapter 3: Running Epik 31
3.1 Running Epik from Maestro 31
3.2 Running Epik from the Command Line 32
3.2.1 Epik Command Line Examples 35
3.2.2 Distributing Epik jobs with para_epik 37
3.3 Epik Limitations 38
Chapter 4: Epik Methodology 39
4.1 pK_a Prediction 39
4.1.1 Overview of Hammett and Taft Prediction of pK_a Values 39
4.1.2 Selection of pK_a^0 and Primary ρ Values 40
4.1.3 ρ Values 41
4.1.4 σ Values for Substituents 43
4.2 Tautomerization 45
4.2.1 Tautomer Pattern Matching 46
4.2.2 Structure Generation 46

4.3 Standard Parameters 47

4.4 Custom Parameters 47

Chapter 5: Structural Adjustment in Epik **49**

5.1 Penalties and Populations for Ionization State and Tautomeric Forms 50

5.2 Creating Structural Variations and Estimating Populations 50

Chapter 6: Getting Help **53**

Appendix A: pK_a Data File Format **55**

A.1 Basic Elements of the Parameter File 55

A.2 Notes about SMARTS patterns for acid_base and special_sigma blocks 60

A.3 Notes about acid_base groups 60

Appendix B: Tautomer Database Format **61**

Index **67**

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Table 3.1.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, and screen output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol | separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Introduction

Epik is a program for the prediction of the pK_a values of the ionizable groups in ligands, and for the generation of the probable ionized and tautomerized structures within a given pH range. Epik rapidly and consistently predicts pK_a values, employing the widely used and respected Hammett and Taft empirical equations. The values of the parameters in these equations are determined from fits to experimental data either from the literature or from Schrödinger's own development efforts. Tautomer probabilities are estimated using a database of tautomerizations derived from experiment and quantum chemical calculations. Both the ionization and tautomerization employ user-modifiable databases of parameters. Collections of probable structures are generated by iteratively tautomerizing and ionizing the base structure.

Since Epik is fast, and yet has a range of options, it is suitable for processing large collections of input structures for a variety of purposes. Epik may also be used by Schrödinger's ligand preparation product, LigPrep, instead of the `ionizer` and `tautomerizer` found in LigPrep. Epik may be run from the command line or more conveniently from Maestro's Epik panel.

1.1 The Epik Manual

This manual explains how to use Epik, a program designed to predict pK_a values of the ionizable groups in ligands, and to generate probable ionized and tautomerized structures within a given pH range. The manual is organized as follows:

- Chapter 1 provides and overview of Epik and the processes it uses to predict pK_a values and generate probable structures.
- Chapter 2 gives a general introduction to Maestro™, the graphical user interface for all Schrödinger's products.
- Chapter 3 explains how to run Epik from Maestro and from the command line.
- Chapter 4 provides detailed explanations of the science behind Epik.
- Chapter 5 details the process of structural adjustment used by Epik.
- Chapter 6 explains how to get help for Epik.
- Appendix A describes the format of the pK_a parameter file.
- Appendix B describes the format of the tautomer database file.

1.2 Epik Technology

This section provides a brief overview of the technology behind Epik. More detailed information is available in Chapter 4. Epik runs in two general modes: pK_a prediction for the structure provided, and protonation/tautomerization state adjustment consistent with a specified pH range. pK_a values are estimated for structures resulting from the latter process.

1.2.1 pK_a Prediction

Epik uses empirical Hammett and Taft relations to predict pK_a values. The first step in this process involves recognizing functional groups that may be ionized by the addition or removal of a proton. Each functional group has a base-line pK_a value and ρ parameter which reflects the sensitivity of the functional group to perturbations from the rest of the molecule. For each such ionizable group the rest of the molecule is conceptually divided into fragments, each of which has a known tendency to perturb ionizable groups. From the tabulated base-line pK_a value, the sensitivity of the functional groups to perturbations, and the strength of the perturbing influences from the various fragments, a prediction of the pK_a for the functional group of this molecule is made.

The various parameters involved in Hammett and Taft equations are derived from fits to experimental data. There is a large collection of such data and pre-fit parameters in the literature. When making pK_a predictions, Epik uses a mixture of literature parameters and parameters determined by Schrödinger. Since Hammett and Taft methodology is empirical it is extremely rapid; typically significantly less than 1 second is needed to estimate the pK_a values for all ionizable sites in a ligand-like molecule when using a 2GHz Pentium 4 processor; and the predicted pK_a values for molecules related to those used in the parametrization are fairly accurate. In addition, Epik attempts to make consistent estimates for the uncertainties of the pK_a values so that the user may elect to seek additional information for uncertain results. Sources of such additional information include experiments on suitable selected model compounds or theoretical estimates of the pK_a values from other programs such as Schrödinger's pK_a Predictor, which is part of the Jaguar suite of programs.

1.2.2 Structure adjustment

In structure adjustment, both the tautomerization and ionization state of the structure may be modified.

Tautomerization is carried out in the same manner, and using the same data, as the `tautomerizer` tool in LigPrep. SMARTS-like patterns from a tautomer database are used to identify and describe how to transform one tautomer into another. Each tautomeric form in the

database is assigned a probability based upon experimental data or, more typically, ab initio quantum mechanical calculations.

In structure adjustment mode Epik adds or removes protons and adjusts the tautomeric state to generate structures that are most probable within a given pH range. We will somewhat loosely refer to this as an ensemble of protonic states consistent with the conditions specified. Since ionization and tautomerization are inter-related the construction of the ensemble is iterative. In each cycle of the iteration all structures accumulated so far are subjected to tautomerization, and then the resulting tautomers are ionized. Structure selection occurs at the end of tautomerization where only the more probable tautomers for each structure are retained; and at the end of the ionization stage where only those species whose overall probability within the currently accumulated ensemble are high enough are retained. When the collection undergoes no change during an iteration the process is judged to be complete. This iteration process can lead to superior results for molecules that can undergo both ionization and tautomerization, particularly if the tautomeric preference varies with the ionization state. At the end of the process the pK_a values are estimated for each functional group present. Each structure present at the end of the iteration process is assigned a probability and a number of penalties based upon: the tautomerizations and ionizations needed to generate it, the collection of molecules generated, and the desired pH. These properties should be useful in penalizing less-probable forms in downstream processing of the structures produced.

Since a number of tautomerizations and ionizations are attempted for a typical ligand-like molecule, and multiple output structures may be produced, the adjustment of the ionization state is considerably more computationally intensive than just estimating the pK_a values for the input structures.

1.3 Installation and Licensing

Epik is licensed separately from other Schrödinger products. While Epik can be used as part of a LigPrep run, it still requires an Epik license when used in this manner.

1.4 Other Utilities

A number of utilities that are provided with Epik, but are not part of Epik itself, might be useful in conjunction with Epik. These utilities are available in `$SCHRODINGER/utilities`, and include:

- `proplister`—extracts requested properties from Maestro-formatted files
- `propfilter`—selects structures from Maestro-formatted files using the property values stored in the files

- `pdbconvert`—converts files between Maestro and PDB formats
- `maesubset`—selects a subset of the structures in a Maestro-formatted file based on structure order
- `sdconvert`—converts files between Maestro and SD formats
- `sdsubset`—selects a subset of the structures in an SD-formatted file

More information on these utilities is available in the Appendix D of the *Maestro User Manual*.

1.5 Citing Epik in Publications

The use of this product should be acknowledged in publications as:

Epik, version 1.0, Schrödinger, LLC, New York, NY, 2005.

Please note that the pK_a and tautomeric databases provided with Epik are copyrighted material, and should not be extracted, reproduced, or used outside of the context of Epik or LigPrep licensed calculations.

Introduction to Maestro

Maestro is the graphical user interface for all of Schrödinger's products: CombiGlide™, Epik™, Glide™, Impact™, Jaguar™, Liaison™, LigPrep™, MacroModel®, Phase™, Prime™, QikProp™, QSite™, SiteMap™, and Strike™. It contains tools for building, displaying, and manipulating chemical structures; for organizing, loading, and storing these structures and associated data; and for setting up, monitoring, and visualizing the results of calculations on these structures. This chapter provides a brief introduction to Maestro and some of its capabilities. For more information on any of the topics in this chapter, see the *Maestro User Manual*.

2.1 General Interface Behavior

Most Maestro panels are amodal: more than one panel can be open at a time, and a panel need not be closed for an action to be carried out. Each Maestro panel has a Close button so you can hide the panel from view.

Maestro supports the mouse functions common to many graphical user interfaces. The left button is used for choosing menu items, clicking buttons, and selecting objects by clicking or dragging. This button is also used for resizing and moving panels. The right button displays a shortcut menu. Other common mouse functions are supported, such as using the mouse in combination with the SHIFT or CTRL keys to select a range of items and select or deselect a single item without affecting other items.

In addition, the mouse buttons are used for special functions described later in this chapter. These functions assume that you have a three-button mouse. If you have a two-button mouse, ensure that it is configured for three-button mouse simulation (the middle mouse button is simulated by pressing or holding down both buttons simultaneously).

2.2 Starting Maestro

Before starting Maestro, you must first set the `SCHRODINGER` environment variable to point to the installation directory. To set this variable, enter the following command at a shell prompt:

csh/tcsh: `setenv SCHRODINGER` *installation-directory*

bash/ksh: `export SCHRODINGER=`*installation-directory*

You might also need to set the DISPLAY environment variable, if it is not set automatically when you log in. To determine if you need to set this variable, enter the command:

```
echo $DISPLAY
```

If the response is a blank line, set the variable by entering the following command:

csh/tcsh: `setenv DISPLAY` *display-machine-name*`:0.0`

bash/ksh: `export DISPLAY=`*display-machine-name*`:0.0`

After you set the SCHRODINGER and DISPLAY environment variables, you can start Maestro using the command:

```
$SCHRODINGER/maestro options
```

If you add the $SCHRODINGER directory to your path, you only need to enter the command maestro. Options for this command are given in Section 2.1 of the *Maestro User Manual*.

The directory from which you started Maestro is Maestro's current working directory, and all data files are written to and read from this directory unless otherwise specified (see Section 2.8 on page 27). You can change directories by entering the following command in the command input area (see page 8) of the main window:

```
cd directory-name
```

where *directory-name* is either a full path or a relative path.

2.3 The Maestro Main Window

The Maestro main window is shown in Figure 2.1 on page 7. The main window components are listed below.

The following components are always visible:

- **Title bar**—displays the Maestro version, the project name (if there is one) and the current working directory.
- **Auto-Help**—automatically displays context-sensitive help.
- **Menu bar**—provides access to panels.
- **Workspace**—displays molecular structures and other 3D graphical objects.

The following components can be displayed or hidden by choosing the component from the Display menu. Your choice of which main window components are displayed is persistent between Maestro sessions.

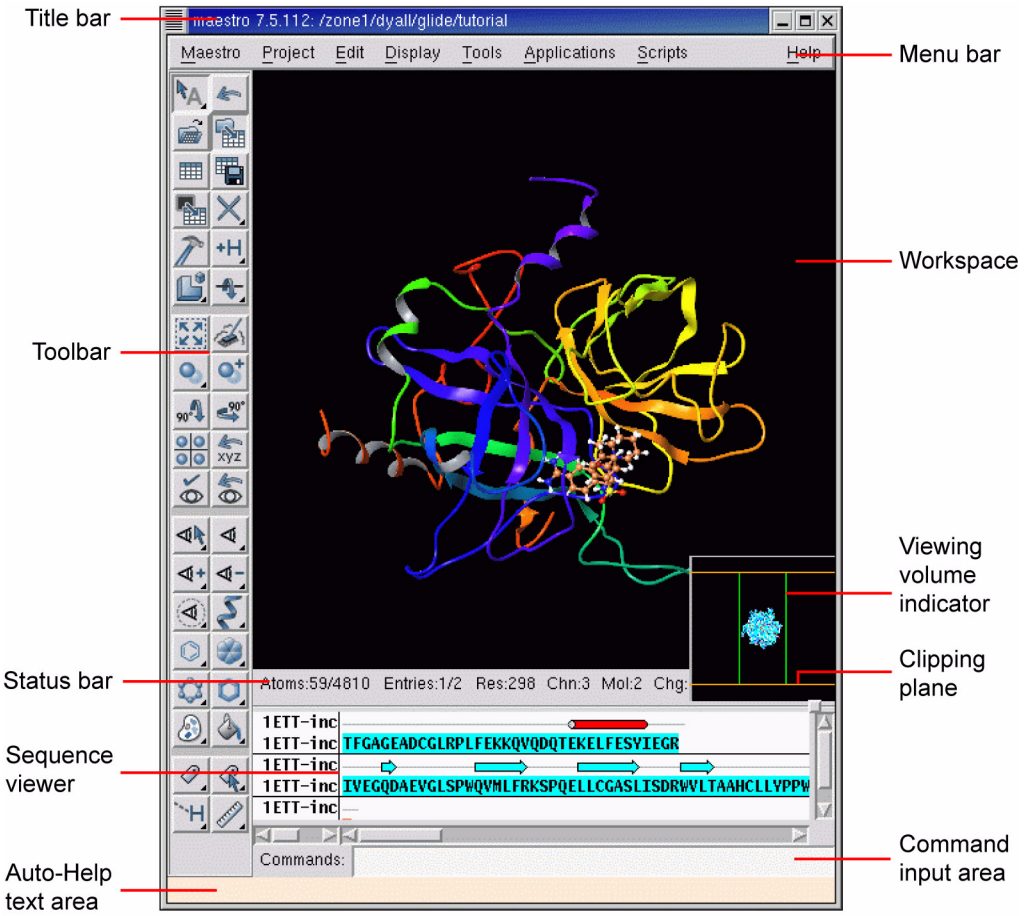

Figure 2.1. The Maestro main window.

- **Toolbar**—contains buttons for many common tasks and provides tools for displaying and manipulating structures, as well as organizing the Workspace.
- **Status bar**—displays information about a particular atom, or about structures in the Workspace, depending on where the pointer pauses (see Section 2.5 of the *Maestro User Manual* for details):
  - **Atom**—displays the chain, residue number, element, PDB atom name, formal charge, and title or entry name (this last field is set by choosing Preferences from the Maestro menu and selecting the Feedback folder).
  - **Workspace**—displays the number of atoms, entries, residues, chains, and molecules in the Workspace.

- **Clipping planes window**—displays a small, top view of the Workspace and shows the clipping planes and viewing volume indicators.
- **Sequence viewer**—shows the sequences for proteins displayed in the Workspace. See Section 2.6 of the *Maestro User Manual* for details.
- **Command input area**—provides a place to enter Maestro commands.

When a distinction between components in the main window and those in other panels is needed, the term *main* is applied to the main window components (e.g., main toolbar).

You can expand the Workspace to occupy the full screen, by pressing CTRL+=. All other components and panels are hidden. To return to the previous display, press CTRL+= again.

2.3.1 The Menu Bar

The menus on the main menu bar provide access to panels, allow you to execute commands, and control the appearance of the Workspace. The main menus are as follows:

- Maestro—save or print images in the Workspace, execute system commands, save or load a panel layout, set preferences, set up Maestro command aliases, and quit Maestro.
- Project—open and close projects, import and export structures, make a snapshot, and annotate a project. These actions can also be performed from the Project Table panel. For more information, see Section 2.4 on page 13.
- Edit—undo actions, build and modify structures, define command scripts and macros, and find atoms in the Workspace.
- Display—control the display of the contents of the Workspace, arrange panels, and display or hide main window components.
- Tools—group atoms; measure, align, and superimpose structures; and view and visualize data.
- Applications—set up, submit, and monitor jobs for Schrödinger's computational programs. Some products have a submenu from which you can choose the task to be performed.
- Scripts—manage and install Python scripts that come with the distribution and scripts that you create yourself. (See Chapter 13 of the *Maestro User Manual* for details.)
- Help—open the Help panel, the PDF documentation index, or information panels; run a demonstration; and display or hide Balloon Help (tooltips).

2.3.2 The Toolbar

The main toolbar contains three kinds of buttons for performing common tasks:

Action—Perform a simple task, like clearing the Workspace.

Display—Open or close a panel or open a dialog box, such as the Project Table panel.

Menu—Display a *button menu*. These buttons have a triangle in the lower right corner.

There are four types of items on button menus, and all four types can be on the same menu (see Figure 2.2):

- **Action**—Perform an action immediately.
- **Display**—Open a panel or dialog box.
- **Object types for selection**—Choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

  The object type is marked on the menu with a red diamond and the button is indented to indicate the action to be performed.
- **Other setting**—Set a state, choose an attribute, or choose a parameter and click on atoms in the Workspace to display or change that parameter.

The toolbar buttons are described below. Some descriptions refer to features not described in this chapter. See the *Maestro User Manual* for a fuller description of these features.

Figure 2.2. The Workspace selection ***button menu and the*** Adjust distances, angles or dihedrals ***button menu.***

Workspace selection
– Choose an object type for selecting
– Open the Atom Selection dialog box

Undo/Redo
Undo or redo the last action. Performs the same function as the Undo item on the Edit menu, and changes to an arrow pointing in the opposite direction when an Undo has been performed, indicating that its next action is Redo.

Open a project
Open the Open Project dialog box.

Import structures
Open the Import panel.

Open/Close Project Table
Open the Project Table panel or close it if it is open.

Save as
Open the Save Project As dialog box, to save the project with a new name.

Create entry from Workspace
Open a dialog box in which you can create an entry in the current project using the contents of the Workspace.

Delete
– Choose an object type for deletion
– Delete hydrogens and waters
– Open the Atom Selection dialog box
– Delete other items associated with the structures in the Workspace
– Click to select atoms to delete
– Double-click to delete all atoms

Open/Close Build panel
Open the Build panel or close it if it is open.

Add hydrogens
– Choose an object type for applying a hydrogen treatment
– Open the Atom Selection dialog box
– Click to select atoms to treat
– Double-click to apply to all atoms

Local transformation
– Choose an object type for transforming
– Click to select atoms to transform
– Open the Advanced Transformations panel

Adjust distances, angles or dihedrals
– Choose a parameter for adjusting
– Delete adjustments

Fit to screen
Scale the displayed structure to fit into the Workspace and reset the center of rotation.

Clear Workspace
Clear all atoms from the Workspace.

Set fog display state
Choose a fog state. Automatic means fog is on when there are more than 40 atoms in the Workspace, otherwise it is off.

Enhance depth cues
Optimize fogging and other depth cues based on what is in the Workspace.

Rotate around X axis by 90 degrees
Rotate the Workspace contents around the X axis by 90 degrees.

Rotate around Y axis by 90 degrees
Rotate the Workspace contents around the Y axis by 90 degrees.

Tile entries
Arrange entries in a rectangular grid in the Workspace.

Reset Workspace
Reset the rotation, translation, and zoom of the Workspace to the default state.

Save view
Save the current view of the Workspace: orientation, location, and zoom.

Restore view
Restore the last saved view of the Workspace: orientation, location, and zoom.

Display only selected atoms
– Choose an object type for displaying
– Click to select atoms to display
– Double-click to display all atoms

Display only
– Choose a predefined atom category
– Open the Atom Selection dialog box

Also display
– Choose a predefined atom category
– Open the Atom Selection dialog box

Undisplay
– Choose a predefined atom category
– Open the Atom Selection dialog box

Display residues within N angstroms of currently displayed atoms
– Choose a radius
– Open a dialog box to set a value

Show, hide, or color ribbons
– Choose to show or hide ribbons
– Choose a color scheme for coloring ribbons

Draw bonds in wire
– Choose an object type for drawing bonds in wire representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw atoms in CPK
– Choose an object type for drawing bonds in CPK representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw atoms in Ball & Stick
– Choose an object type for drawing bonds in Ball & Stick representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw bonds in tube
– Choose an object type for drawing bonds in tube representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Color all atoms by scheme
Choose a predefined color scheme.

Color residue by constant color
– Choose a color for applying to residues
– Click to select residues to color
– Double-click to color all atoms

Label atoms
– Choose a predefined label type
– Delete labels

Label picked atoms
– Choose an object type for labeling atoms
– Open the Atom Selection dialog box
– Open the Atom Labels panel at the Composition folder
– Delete labels
– Click to select atoms to label
– Double-click to label all atoms

Display H-bonds
– Choose bond type:
intra—displays H-bonds within the selected molecule
inter—displays H-bonds between the selected molecule and all other atoms.
– Delete H-bonds
– Click to select molecule

Measure distances, angles or dihedrals
– Choose a parameter for displaying measurements
– Delete measurements
– Click to select atoms for measurement

2.3.3 Mouse Functions in the Workspace

The left mouse button is used for selecting objects. You can either click on a single atom or bond, or you can drag to select multiple objects. The right mouse button opens shortcut menus, which are described in Section 2.7 of the *Maestro User Manual*.

The middle and right mouse buttons can be used on their own and in combination with the SHIFT and CTRL keys to perform common operations, such as rotating, translating, centering, adjusting, and zooming.

Table 2.1. Mapping of Workspace operations to mouse actions.

Mouse Button	Keyboard	Motion	Action
Left		click, drag	Select
Left	SHIFT	click, drag	Toggle the selection
Middle		drag	Rotate about X and Y axes Adjust bond, angle, or dihedral
Middle	SHIFT	drag vertically	Rotate about X axis
Middle	SHIFT	drag horizontally	Rotate about Y axis
Middle	CTRL	drag horizontally	Rotate about Z axis
Middle	SHIFT + CTRL	drag horizontally	Zoom
Right		click	Spot-center on selection
Right		click and hold	Display shortcut menu
Right		drag	Translate in the X-Y plane
Right	SHIFT	drag vertically	Translate along the X axis
Right	SHIFT	drag horizontally	Translate along the Y axis
Right	CTRL	drag horizontally	Translate along the Z axis
Middle & Right		drag horizontally	Zoom

2.3.4 Shortcut Key Combinations

Some frequently used operations have been assigned shortcut key combinations. The shortcuts available in the main window are described in Table 2.2.

Table 2.2. Shortcut keys in the Maestro main window.

Keys	Action	Equivalent Menu Choices
CTRL+B	Open Build panel	Edit > Build
CTRL+C	Create entry	Project > Create Entry From Workspace
CTRL+E	Open Command Script Editor panel	Edit > Command Script Editor
CTRL+F	Open Find Atoms panel	Edit > Find
CTRL+H	Open Help panel	Help > Help
CTRL+I	Open Import panel	Project > Import Structures
CTRL+M	Open Measurements panel	Tools > Measurements
CTRL+N	Create new project	Project > New
CTRL+O	Open project	Project > Open
CTRL+P	Print	Maestro > Print
CTRL+Q	Quit	Maestro > Quit
CTRL+S	Open Sets panel	Tools > Sets
CTRL+T	Open Project Table panel	Project > Show Table
CTRL+W	Close project	Project > Close
CTRL+Z	Undo/Redo last command	Edit > Undo/Redo
CTRL+=	Enter and exit full screen mode (Workspace occupies full screen)	None

2.4 Maestro Projects

All the work you do in Maestro is done within a *project*. A project consists of a set of *entries*, each of which contains one or more chemical structures and their associated data. In any Maestro session, there can be only one Maestro project open. If you do not specify a project when you start Maestro, a *scratch* project is created. You can work in a scratch project without saving it, but you must save it in order to use it in future sessions. When you save or close a project, all the view transformations (rotation, translation, and zoom) are saved with it. When you close a project, a new scratch project is automatically created.

Likewise, if there is no entry displayed in the Workspace, Maestro creates a *scratch* entry. Structures that you build in the Workspace constitute a scratch entry until you save the structures as project entries. The scratch entry is not saved with the project unless you explicitly add it to the project. However, you can use a scratch entry as input for some calculations.

To add a scratch entry to a project, do one of the following:

- Click the Create entry from Workspace button:

- Choose Create Entry from Workspace from the Project menu.
- Press CTRL+C.

In the dialog box, enter a name and a title for the entry. The entry name is used internally to identify the entry and can be modified by Maestro. The title can be set or changed by the user, but is not otherwise modified by Maestro.

Once an entry has been incorporated into the project, its structures and their data are represented by a row in the Project Table. Each row contains the row number, an icon indicating whether the entry is displayed in the Workspace (the In column), the entry title, a button to open the Surfaces panel if the entry has surfaces, the entry name, and any entry properties. The row number is not a property of the entry.

Entries can be collected into groups, and the members of the group can be displayed or hidden. Most additions of multiple entries to the Project Table are done as entry groups.

You can use entries as input for all of the computational programs—Glide, Impact, Jaguar, Liaison, LigPrep, MacroModel, Phase, Prime, QikProp, QSite, and Strike. You can select entries as input for the ePlayer, which displays the selected structures in sequence. You can also duplicate, combine, rename, and sort entries; create properties; import structures as entries; and export structures and properties from entries in various formats.

To open the Project Table panel, do one of the following:

- Click the Open/Close Project Table button on the toolbar

- Choose Show Table from the Project menu
- Press CTRL+T.

The Project Table panel contains a menu bar, a toolbar, and the table itself.

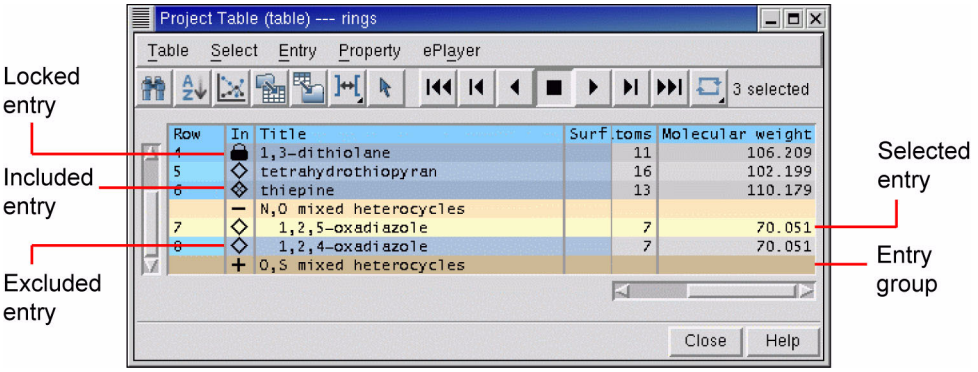

Figure 2.3. The** Project Table **panel.

2.4.1 The Project Table Toolbar

The Project Table toolbar contains two groups of buttons and a status display. The first set of buttons opens various panels that allow you to perform functions on the entries in the Project Table. The second set of buttons controls the ePlayer, which "plays through" the selected structures: each structure is displayed in the Workspace in sequence, at a given time interval. See Section 2.3.2 on page 9 for a description of the types of toolbar buttons. The buttons are described below.

Find
Open the Find panel for locating alphanumeric text in any column of the Project Table, except for the row number.

Sort
Open the Sort panel for sorting entries by up to three properties.

Plot
Open the Plot panel for plotting entry properties.

Import Structure
Open the Import panel for importing structures into the project.

Export Structure
Open the Export panel for exporting structures to a file.

Columns
Choose an option for adjusting the column widths.

Select only
Open the Entry Selection dialog box for selecting entries based on criteria for entry properties.

Go to start
Display the first selected structure.

Previous
Display the previous structure in the list of selected structures.

Play backward
Display the selected structures in sequence, moving toward the first.

Stop
Stop the ePlayer.

Play forward
Display the selected structures in sequence, moving toward the last.

Next
Display the next structure in the list of selected structures.

Go to end
Display the last selected structure.

Loop
Choose an option for repeating the display of the structures. Single Direction displays structures in a single direction, then repeats. Oscillate reverses direction each time the beginning or end of the list is reached.

The status display, to the right of the toolbar buttons, shows the number of selected entries. When you pause the cursor over the status display, the Balloon Help shows the total number of entries, the number shown in the table, the number selected, and the number included in the Workspace.

2.4.2 The Project Table Menus

- Table—find text, sort entries, plot properties, import and export structures, and configure the Project Table.
- Select—select all entries, none, invert your selection, or select classes of entries using the Entry Selection dialog box and the Filter panel.
- Entry—include or exclude entries from the Workspace, display or hide entries in the Project Table, and perform various operations on the selected entries.
- Property—display and manipulate entry properties in the Project Table.
- ePlayer—view entries in succession, stop, reverse, and set the ePlayer options.

2.4.3 Selecting Entries

Many operations in Maestro are performed on the entries selected in the Project Table. The Project Table functions much like any other table: select rows by clicking, shift-clicking, and control-clicking. However, because clicking in an editable cell of a selected row enters edit mode, you should click in the Row column to select entries. See Section 2.4.5 on page 18 for more information on mouse actions in the Project Table. There are shortcuts for selecting classes of entries on the Select menu.

In addition to selecting entries manually, you can select entries that meet a combination of conditions on their properties. Such combinations of conditions are called *filters*. Filters are Entry Selection Language (ESL) expressions and are evaluated at the time they are applied. For example, if you want to set up a Glide job that uses ligands with a low molecular weight (say, less than 300) and that has certain QikProp properties, you can set up a filter and use it to select entries for the job. If you save the filter, you can use it again on a different set of ligands that meet the same selection criteria.

To create a filter:

1. Do one of the following:
   - Choose Only, Add, or Deselect from the Select menu.
   - Click the Entry selection button on the toolbar.

   
2. In the Properties folder, select a property from the property list, then select a condition.
3. Combine this selection with the current filter by clicking Add, Subtract, or Intersect. These buttons perform the Boolean operations OR, AND NOT, and AND on the corresponding ESL expressions.
4. To save the filter for future use click Create Filter, enter a name, and click OK.
5. Click OK to apply the filter immediately.

2.4.4 Including Entries in the Workspace

In addition to selecting entries, you can also use the Project Table to control which entries are displayed in the Workspace. An entry that is displayed in the Workspace is *included* in the Workspace; likewise, an entry that is not displayed is *excluded*. Included entries are marked by an X in the diamond in the In column; excluded entries are marked by an empty diamond. Entry inclusion is completely independent of entry selection.

To include or exclude entries, click, shift-click, or control-click in the In column of the entries, or select entries and choose Include or Exclude from the Entry menu. Inclusion with the mouse works just like selection: when you include an entry by clicking, all other entries are excluded.

It is sometimes useful to keep one entry in the Workspace and include others one by one: for example, a receptor and a set of ligands. You can fix the receptor in the Workspace by selecting it in the Project Table and choosing Fix from the Entry menu or by pressing CTRL+F. A padlock icon replaces the diamond in the In column to denote a *fixed* entry. To remove a fixed entry from the Workspace, you must exclude it explicitly (CTRL+X). It is not affected by the inclusion or exclusion of other entries. Fixing an entry affects only its inclusion; you can still rotate, translate, or modify the structure.

2.4.5 Mouse Functions in the Project Table

The Project Table supports the standard use of shift-click and control-click to select objects. This behavior applies to the selection of entries and the inclusion of entries in the Workspace. You can also drag to resize rows and columns and to move rows.

You can drag a set of non-contiguous entries to reposition them in the Project Table. When you release the mouse button, the entries are placed after the first unselected entry that precedes the entry on which the cursor is resting. For example, if you select entries 2, 4, and 6, and release the mouse button on entry 3, these three entries are placed after entry 1, because entry 1 is the first unselected entry that precedes entry 3. To move entries to the top of the table, drag them above the top of the table; to move entries to the end of the table, drag them below the end of the table.

A summary of mouse functions in the Project Table is provided in Table 2.3.

Table 2.3. Mouse operations in the Project Table.

Task	Mouse Operation
Change a Boolean property value	Click repeatedly in a cell to cycle through the possible values (On, Off, Clear)
Display the Entry menu for an entry	Right-click anywhere in the entry. If the entry is not selected, it becomes the selected entry. If the entry is selected, the action is applied to all selected entries.
Display a version of the Property menu for a property	Right-click in the column header
Edit the text or the value in a table cell	Click in the cell and edit the text or value
Include an entry in the Workspace, exclude all others	Click the In column of the entry

Table 2.3. Mouse operations in the Project Table. (Continued)

Task	Mouse Operation
Move selected entries	Drag the entries
Paste text into a table cell	Middle-click
Resize rows or columns	Drag the boundary with the middle mouse button
Select an entry, deselect all others	For an unselected entry, click anywhere in the row except the In column; for a selected entry, click the row number.
Select or include multiple entries	Click the first entry then shift-click the last entry
Toggle the selection or inclusion state	Control-click the entry or the In column

2.4.6 Project Table Shortcut Keys

Some frequently used project operations have been assigned shortcut key combinations. The shortcuts, their functions, and their menu equivalents are listed in Table 2.4.

Table 2.4. Shortcut keys in the Project Table.

Keys	Action	Equivalent Menu Choices
CTRL+A	Select all entries	Select > All
CTRL+F	Fix entry in Workspace	Entry > Fix
CTRL+I	Open Import panel	Table > Import Structures
CTRL+N	Include only selected entries	Entry > Include Only
CTRL+U	Deselect all entries	Select > None
CTRL+X	Exclude selected entries	Entry > Exclude
CTRL+Z	Undo/Redo last command	Edit > Undo/Redo in main window

2.5 Building a Structure

After you start Maestro, the first task is usually to create or import a structure. You can open existing Maestro projects or import structures from other sources to obtain a structure, or you can build your own. To open the Build panel, do one of the following:

- Click the Open/Close Build panel button in the toolbar:

- Choose Build from the Edit menu.
- Press CTRL+B.

The Build panel allows you to create structures by drawing or placing atoms or fragments in the Workspace and connecting them into a larger structure, to adjust atom positions and bond orders, and to change atom properties. This panel contains a toolbar and three folders.

2.5.1 Placing and Connecting Fragments

The Build panel provides several tools for creating structures in the Workspace. You can place and connect fragments, or you can draw a structure freehand.

To place a fragment in the Workspace:

1. Select Place.
2. Choose a fragment library from the Fragments menu.
3. Click a fragment.
4. Click in the Workspace where you want the fragment to be placed.

To connect fragments in the Workspace, do one of the following:

- Place another fragment and connect them using the Connect & Fuse panel, which you open from the Edit menu on the main menu bar or with the Display Connect & Fuse panel on the Build toolbar.

- Replace one or more atoms in the existing fragment with another fragment by selecting a fragment and clicking in the Workspace on the main atom to be replaced.
- Grow another fragment by selecting Grow in the Build panel and clicking the fragment you want to add in the Fragments folder.

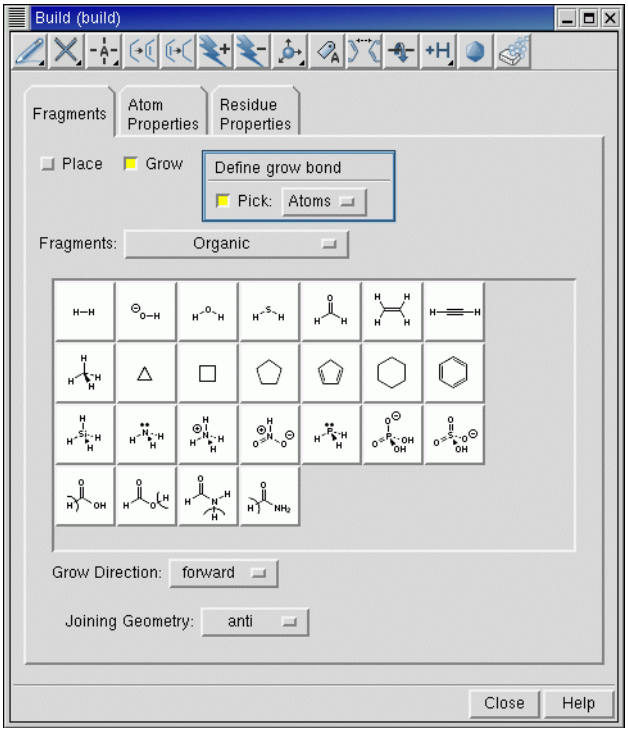

Figure 2.4. The* Build *panel.

Grow mode uses predefined rules to connect a fragment to the *grow bond*. The grow bond is marked by a green arrow. The new fragment replaces the atom at the head of the arrow on the grow bond and all atoms attached to it. To change the grow bond, choose Bonds from the Pick option menu in the Build panel and click on the desired grow bond in the Workspace. The arrow points to the atom nearest to where you clicked.

To draw a structure freehand:

1. Choose an element from the Draw button menu on the Build panel toolbar:

   

2. Click in the Workspace to place an atom of that element.
3. Click again to place another atom and connect it to the previous atom.
4. Continue this process until you have drawn the structure.
5. Click the active atom again to finish drawing.

2.5.2 Adjusting Properties

In the Atom Properties folder, you can change the properties of the atoms in the Workspace. For each item on the Property option menu—Element, Atom Type (MacroModel), Partial Charge, PDB Atom Name, Grow Name, and Atom Name—there is a set of tools you can use to change the atom properties. For example, the Element tools consist of a periodic table from which you can choose an element and select an atom to change it to an atom of the selected element.

Similarly, the Residue Properties folder provides tools for changing the properties of residues: the Residue Number, the Residue Name, and the Chain Name.

To adjust bond lengths, bond angles, dihedral angles, and chiralities during or after building a structure, use the Adjust distances, angles or dihedrals button on the main toolbar:

You can also open the Adjust panel from this button menu, from the Display Adjust panel button on the Build panel toolbar (which has the same appearance as the above button) or from the Edit menu in the main window.

2.5.3 The Build Panel Toolbar

The toolbar of the Build panel provides quick access to tools for drawing and modifying structures and labeling atoms. See Section 2.3.2 on page 9 for a description of the types of toolbar buttons. The toolbar buttons and their use are described below.

Free-hand drawing
Choose an element for drawing structures freehand in the Workspace (default C). Each click in the Workspace places an atom and connects it to the previous atom.

Delete
Choose an object for deleting. Same as the Delete button on the main toolbar, see page 10.

Set element
Choose an element for changing atoms in the Workspace (default C). Click an atom to change it to the selected element.

Increment bond order
Select a bond to increase its bond order by one, to a maximum of 3.

Decrement bond order
Select a bond to decrease its bond order by one, to a minimum of 0.

Increment formal charge
Select an atom to increase its formal charge by one.

Decrement formal charge
Select an atom to decrease its formal charge by one.

Move
Choose a direction for moving atoms, then click the atom to be moved. Moves in the XY plane are made by clicking the new location. Moves in the Z direction are made in 0.5 Å increments.

Label
Apply heteroatom labels as you build a structure. The label consists of the element name and formal charge, and is applied to atoms other than C and H.

Display Connect & Fuse panel
Open the Connect & Fuse panel so you can connect structures (create bonds between structures) or fuse structures (replace atoms of one structure with those of another).

Display Adjust panel
Open the Adjust panel so you can change bond lengths, bond angles, dihedral angles, or atom chiralities.

Add hydrogens
Choose an atom type for applying the current hydrogen treatment. Same as the Add hydrogens button on the main toolbar, see page 10.

Geometry Symmetrizer
Open the Geometry Symmetrizer panel for symmetrizing the geometry of the structure in the Workspace.

Geometry Cleanup
Clean up the geometry of the structure in the Workspace.

2.6 Selecting Atoms

Maestro has a powerful set of tools for selecting atoms in a structure: toolbar buttons, picking tools in panels, and the Atom Selection dialog box. These tools allow you to select atoms in two ways:

- Select atoms first and apply an action to them
- Choose an action first and then select atoms for that action

2.6.1 Toolbar Buttons

The small triangle in the lower right corner of a toolbar button indicates that the button contains a menu. Many of these buttons allow you to choose an object type for selecting: choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

For example, to select atoms with the Workspace selection toolbar button:

1. Choose Residues from the Workspace selection button menu:

   

   The button changes to:

   

2. Click on an atom in a residue in the Workspace to select all the atoms in that residue.

2.6.2 Picking Tools

The picking tools are embedded in each panel in which you need to select atoms to apply an operation. The picking tools in a panel can include one or more of the following:

- Pick option menu—Allows you to choose an object type. Depending on the operation to be performed, you can choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

  The Pick option menu varies from panel to panel, because not all object types are appropriate for a given operation. For example, some panels have only Atoms and Bonds in the Pick option menu.

- All button—Performs the action on all atoms in the Workspace.
- Selection button—Performs the action on any atoms already selected in the Workspace.
- Previous button—Performs the action on the most recent atom selection defined in the Atom Selection dialog box.
- Select button—Opens the Atom Selection dialog box.
- ASL text box—Allows you to type in an ASL expression for selecting atoms.

  ASL stands for Atom Specification Language, and is described in detail in the *Maestro Command Reference Manual*.

- Clear button—Clears the current selection

  

- Show markers option—Marks the selected atoms in the Workspace.

For example, to label atoms with the Label Atoms panel:

1. Choose Atom Labels from the Display menu.
2. In the Composition folder, select Element and Atom Number.
3. In the picking tools section at the top of the panel, you could do one of the following:
   - Click Selection to apply labels to the atoms already selected in the Workspace (from the previous example).
   - Choose Residues from the Pick option menu and click on an atom in a different residue to label all the atoms in that residue.

2.6.3 The Atom Selection Dialog Box

If you wish to select atoms based on more complex criteria, you can use the Atom Selection dialog box. To open this dialog box, choose Select from a button menu or click the Select button in a panel. See Section 5.3 of the *Maestro User Manual* for detailed instructions on how to use the Atom Selection dialog box.

2.7 Scripting in Maestro

Although you can perform nearly all Maestro-supported operations through menus and panels, you can also perform operations using Maestro commands, or compilations of these commands, called *scripts*. Scripts can be used to automate lengthy procedures or repetitive tasks and can be created in several ways. These are summarized below.

2.7.1 Python Scripts

Python is a full-featured scripting language that has been embedded in Maestro to extend its scripting facilities. The Python capabilities within Maestro include access to Maestro functionality for dealing with chemical structures, projects, and Maestro files.

The two main Python commands used in Maestro are:

- `pythonrun`—executes a Python module. (You can also use the alias `pyrun`.) The syntax is:

  `pythonrun` *module*`.`*function*

- `pythonimport`—rereads a Python file so that the next time you use the `pythonrun` command, it uses the updated version of the module. (You can also use the alias `pyimp`.)

From the Maestro Scripts menu you can install, manage, and run Python scripts. For more information on the Scripts menu, see Section 13.1 of the *Maestro User Manual*.

For more information on using Python with Maestro, see *Scripting with Python*.

2.7.2 Command Scripts

All Maestro commands are logged and displayed in the Command Script Editor panel. This means you can create a command script by performing the operations with the GUI controls, copying the logged commands from the Command History list into the Script text area of the panel, then saving the list of copied commands as a script.

To run an existing command script:

1. Open the Command Script Editor panel from the Edit menu in the main window.
2. Click Open Local and navigate to the directory containing the desired script.
3. Select a script in the Files list and click Open.

   The script is loaded into the Script window of the Command Script Editor panel.
4. Click Run Script.

Command scripts cannot be used for Prime operations.

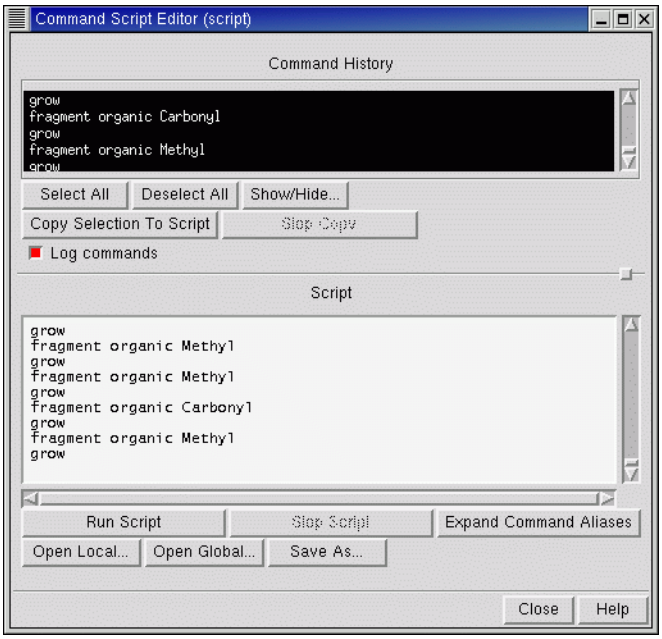

Figure 2.5. The Command Script Editor ***panel.***

2.7.3 Macros

There are two kinds of macros you can create: named macros and macros assigned to function keys F1 through F12.

To create and run a named macro:

1. Open the Macros panel from the Edit menu in the main window.
2. Click New, enter a name for the macro, and click OK.
3. In the Definition text box, type the commands for the macro.
4. Click Update to update the macro definition.
5. To run the macro, enter the following in the command input area in the main window:

   `macrorun` *macro-name*

   If the command input area is not visible, choose Command Input Area from the Display menu.

To create and run a function key macro:

1. Open the Function Key Macros panel from the Edit menu in the main window.
2. From the Macro Key option, select a function key (F1 through F12) to which to assign the macro.
3. In the text box, type the commands for the macro.
4. Click Run to test the macro or click Save to save it.
5. To run the macro from the main window, press the assigned function key.

For more information on macros, see Section 13.5 of the *Maestro User Manual*.

2.8 Specifying a Maestro Working Directory

When you use Maestro to launch Epik jobs, Maestro writes job output to the directory specified in the Directory folder of the Preferences panel. By default, this directory (the file I/O directory) is the directory from which you started Maestro.

To change the Maestro working directory:

1. Open the Preferences panel from the Maestro menu.
2. Click the Directory tab.
3. Select the directory you want to use for reading and writing files.

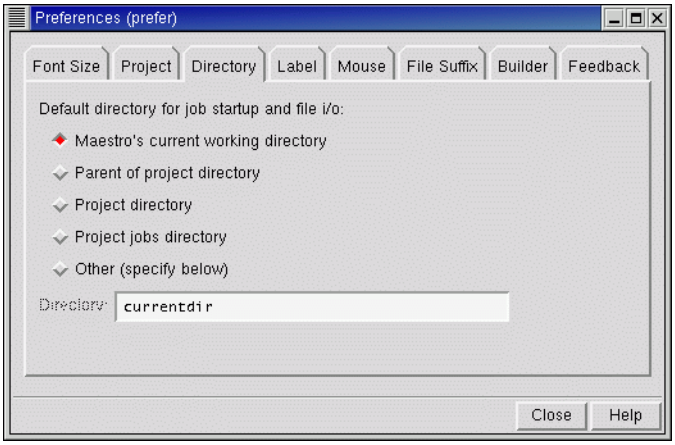

Figure 2.6. The** Directory **folder of the** Preferences **panel.

You can also set other preferences in the Preferences panel. See Section 12.2 of the *Maestro User Manual* for details.

2.9 Undoing an Operation

To undo a single operation, click the Undo button in the toolbar, choose Undo from the Edit menu, or press CTRL+Z. The word Undo in the menu is followed by text that describes the operation to undo. Not all operations can be undone: for example, global rotations and translations are not undoable operations. For such operations you can use the Save view and Restore view buttons in the toolbar, which save and restore a molecular orientation.

2.10 Running and Monitoring Jobs

Maestro has panels for each product for preparing and submitting jobs. To use these panels, choose the appropriate product and task from the Applications menu and its submenus. Set the appropriate options in the panel, then click Start to open the Start dialog box and set options for running the job. For a complete description of the Start dialog box associated with your computational program, see your product's User Manual. When you have finished setting the options, click Start to launch the job and open the Monitor panel.

The Monitor panel is the control panel for monitoring the progress of jobs and for pausing, resuming, or killing jobs. All jobs that belong to you can be displayed in the Monitor panel, whether or not they were started from Maestro. Subjobs are indented under their parent in the job list. The text pane shows output information from the monitored job, such as the contents

of the log file. The Monitor panel opens automatically when you start a job. If it is not open, you can open it by choosing Monitor from the Applications menu in the Maestro main window.

While jobs are running, the Detach, Pause, Resume, Stop, Kill, and Update buttons are active. When there are no jobs currently running, only the Monitor and Delete buttons are active. These buttons act on the selected job. By default, only jobs started from the current project are shown. To show other jobs, deselect Show jobs from current project only.

When a monitored job ends, the results are incorporated into the project according to the settings used to launch the job. If a job that is not currently being monitored ends, you can select it in the Monitor panel and click Monitor to incorporate the results. Monitored jobs are incorporated only if they are part of the current project. You can monitor jobs that are not part of the current project, but their results are not incorporated. To add their results to a project, you must open the project and import the results.

Further information on job control, including configuring your site, monitoring jobs, running jobs, and job incorporation, can be found in the *Job Control Guide* and the *Installation Guide*.

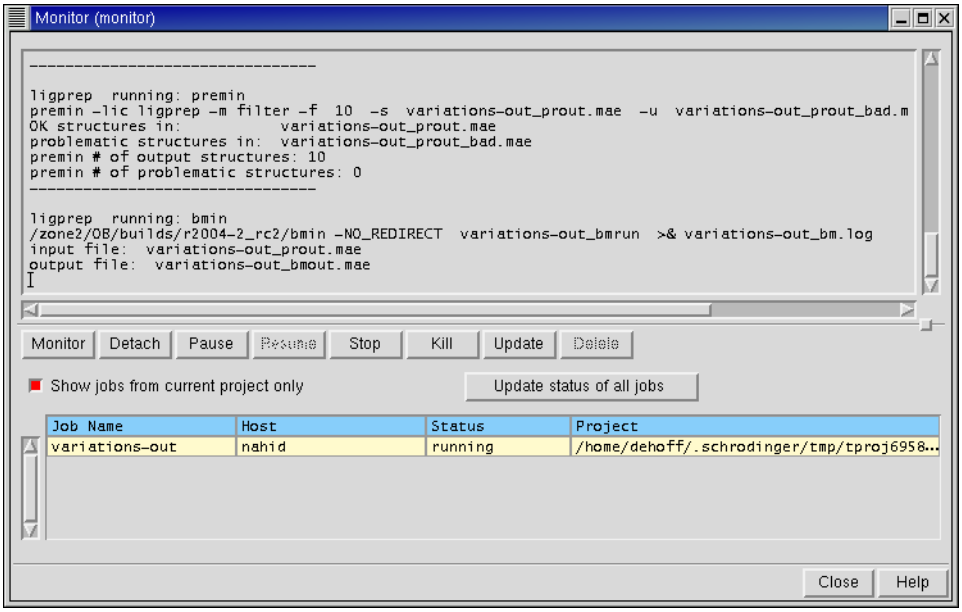

Figure 2.7. The* Monitor *panel.

2.11 Getting Help

Maestro comes with automatic, context-sensitive help (Auto-Help), Balloon Help (tooltips), an online help facility, and a user manual. To get help, follow the steps below:

- Check the Auto-Help text box at the bottom of the main window. If help is available for the task you are performing, it is automatically displayed there. It describes what actions are needed to perform the task.
- If your question concerns a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- If you do not find the help you need using either of the steps above, click the Help button in the lower right corner of the appropriate panel. The Help panel is displayed with a relevant help topic.
- For help with a concept or action not associated with a panel, open the Help panel from the Help menu or press CTRL+H.

If you do not find the information you need in the Maestro help system, check the following sources:

- The *Maestro User Manual*
- The Frequently Asked Questions page on the Schrödinger Support Center.

You can also contact Schrödinger by e-mail or phone for help:

- E-mail: help@schrodinger.com
- Phone: (503) 299-1150

2.12 Ending a Maestro Session

To end a Maestro session, choose Quit from the Maestro menu. To save a log file with a record of all operations performed in the current session, click Quit, save log file in the Quit panel. This information can be useful to Schrödinger support staff when responding to any problem you report.

Running Epik

3.1 Running Epik from Maestro

Epik jobs can be submitted from the Epik panel in Maestro. To open the Epik panel, choose Epik from the Applications menu in the main window.

The two tasks available from the Epik panel are the prediction of the pK_a values of the ionizable atoms in a set of structures, and the generation of the probable ionized (and tautomerized) states of a set of structures within a given pH range. The pK_a predictions are rule-based, and so can be generated very rapidly. The pK_a values and their uncertainties are stored as atomic properties with the structure.

When the Epik job finishes, the pK_a values are automatically displayed on the structures in the Workspace as atom labels. Previously calculated pK_a values can be viewed using the Atom Labels panel. The labels can also be cleared in the Atom Labels panel.

To predict the pKa of existing structures:

1. Select the source of the structures using the tools at the top of the panel.
2. Select Query only for the Analysis mode.
3. Choose the solvent from the Solvent option menu.
4. Click Start.
5. Set job parameters in the Start dialog box, and click Start.

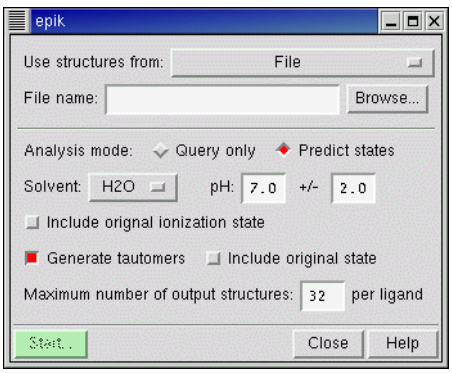

Figure 3.1. The** Epik **panel.

To generate the ionization states of structures within a given pH range:

1. Select the source of the structures using the tools at the top of the panel.
2. Select Predict states for the Analysis mode.
3. Choose the solvent from the Solvent option menu.
4. Enter the target pH and range in the pH text boxes.

   The range is converted into a probability of 10^{-range}. Ionized (and tautomerized) structures whose probability exceeds this value, when all likely structures are considered, are kept. If the structure does not tautomerize, this is equivalent to keeping structures whose groups are ionized if their pK_a lies within the specified range of the target pH value.
5. If you want to keep the original ionization state, regardless of its probability in the given pH range, select Include original ionization state.
6. Select Generate tautomers to tautomerize structures during the structural adjustment process.
7. If you selected Generate tautomers and want to keep the original tautomer, regardless of its probability in the given pH range, select Include original tautomer.
8. Enter the maximum number of structures to generate for each input structure in the Maximum number of output structures text box.
9. Click Start.
10. Set job parameters in the Start dialog box, and click Start.

3.2 Running Epik from the Command Line

Epik can be run using the `epik` command. The syntax for this command is shown below.

Syntax:

`epik [`*options*`] -imae` *input_file*`.mae -omae` *output_file*`.mae`

The input and output files are required and must be specified using the `-imae` and `-omae` options respectively. Only Maestro-formatted files are supported. The input file can contain more than one structure.

The options for the `epik` commands are listed in Table 3.1, which can be displayed using the `-h` option. The job level and diagnostic options, most of which are common to Schrödinger products, are given in Table 3.2 and Table 3.3. The diagnostic options provide information, but the program does not run.

Table 3.1. Options for the `epik` *command*

Option	Description
`-a`	Estimate pK_a values for acids only.
`-b`	Estimate pK_a values for bases only.
`-es` *filename*	Use the specified custom pKa parameter file. Default is `pKa_water_HT_data`.
`-h[elp]|` `-HELP`	Print usage message.
`-ma` *number*	Structures containing more than *number* atoms will not be adjusted. Default is 150.
`-ms` *number*	Maximum number of generated structures per input structure. Default is 32.
`-nt`	Do not tautomerize structures.
`-p` *value*	The minimum probability at the target pH for generated states to be kept. Default is 0.01. Probabilities are evaluated on the basis of an ensemble of likely states, as determined from ionization and tautomerization equilibria.
`-ph` *value*	Target pH for generated states. No default.
`-pht` *value*	pH tolerance for generated structures. The minimum probability for generated states is determined from $-\log_{10}(p) = value$. When tautomerization is disallowed, this is equivalent to keeping structures whose pK_a value lies within *value* units of the target pH value. No default.
`-pKa_atom`	Report pK_a of specific atoms as a Maestro property.
`-retain_i`	Retain the initial ionization state. Also retains the initial tautomerization state.
`-retain_i_lab`	Similar to `-retain_i` except that only labeled input structures are retained.
`-retain_t`	Retain the initial tautomerization state.
`-retain_t_lab`	Similar to `-retain_t` except that only labeled input structures are retained.
`-tn` *value*	Maximum number of tautomers. Default is 8.
`-tp` *value*	Minimum probability for tautomers.
`-ts` *filename*	Use the specified custom tautomer database file.
`-v`	Print the version number.
`-verb` *number*	controls MMERR level of reporting. 0 `MMERR_FATAL` - only report errors 1 `MMERR_WARNING` - only report errors and warnings 2 `MMERR_INFO` - report errors, warnings, and processing information. 3 `MMERR_DEBUG` - report errors, warnings, processing information, and debugging information

Table 3.2. Job control options for the `epik` *command.*

Option	Description
`-DEBUG`	Show details of operation of the top level scripts.
`-HOST` *hostname*	The name of the host to use for this run. The host name must be defined in the `schrodinger.hosts` file.
`-HOSTFILE` *filename*	The name of the host file to use for this run. The default host file is the installed version of `schrodinger.hosts`.
`-INTERVAL`	The maximum time in seconds between updates of the *jobname*`.log` file.
`-LOCAL`	Do not use a temporary directory for intermediate files. Keep files in the current working directory.
`-NICE`	Run the job at reduced priority.
`-NO_JOBCONTROL`	Do not use job control; print Epik messages to *jobname*`.log`.
`-NO_REDIRECT`	Do not use job control; print Epik messages to the terminal window.
`-PROJ` *name*	Assign job to the Maestro project *name*.
`-TMPDIR` *directory*	The name of the directory used to store files temporarily during a job.
`-USER` *name*	Launch job as user *name*.
`-WAIT`	Do not return a prompt until the job finishes.

Table 3.3. Diagnostic options for the `epik` *command*

Option	Description
`-ALL`	Ignore platform compatibility (list all platforms).
`-ENTRY`	List the relevant host entries from the hosts file.
`-HOSTS`	List the hosts available for remote jobs.
`-LIST`	List all the platform-compatible versions of product.
`-WHICH`	Report the product exec directory and `MMSHARE_EXEC`.
`-WHY`	Show details of how the selected executable directory was chosen.

Epik jobs run in the background under Schrödinger's Job Control facility unless the `-NO_JOBCONTROL` or `-NO_REDIRECT` options are given. Job Control runs the `epik` command in a temporary directory where up-to-date files are maintained while the job is running. When the job finishes, the output files are copied back to the job submission directory. While the job is running the `.log` file is monitored. That is, an up-to-date copy of this file is maintained in the job submission directory. Specifying the `-LOCAL` option overrides the use of a temporary

directory and an attempt is made to run out of the job submission directory. For more information on the Job Control facility, see the *Job Control Guide*.

The `-ph` Option

- Unless `-ph` is specified, pK_a values are estimated using the input structure.
- When `-ph` is specified, variations on the input structure are generated, and the pK_a values for each of them is calculated. If no structure meets the criteria for output the most probable structure is kept so that at least one structure is always produced for each input structure.

Notes on Retention Behavior

- If you forcibly retain structures, they are selected for inclusion in the output ahead of any more probable structures.
- The option `-retain_i` forces the retention of the exact combination of the input tautomerization and ionization state.
- The option `-retain_t` forces retention of the various ionized forms of the input tautomerization state based upon their ionization probabilities. At a minimum the most probable ionization state of the input tautomeric form is retained.
- If you specify both `-retain_i` and `-retain_t`, a combination of the behaviors occurs. A copy of the exact combination of the input tautomerization and ionization state plus probable ionization states of the input tautomer are retained. As well, variations on the ionization state for the input tautomer will be prioritized based upon the ionization penalties only.

3.2.1 Epik Command Line Examples

The Epik panel in Maestro is a convenient and general tool for running Epik, particularly for small batches of molecules. Epik can also be run from the command line, and thus run from user-written scripts.

- The default behavior of Epik is to find pK_a values for the input structures. To run this process from the command line, only the input and output file names need to be specified:

  ```
  $SCHRODINGER/epik -imae ligands.mae -omae ligands_with_pKas.mae
  ```

  The output structure file will contain the same structures that were present in the input file, with atom level properties for the pK_a values. For acids, the pK_a values are associated with the acid hydrogens; while for bases, the pK_a values are associated with the atoms to which the proton would bond. Note that the pK_a values are for the input structure as given, regardless of how suitable the input structures actually are.

- To adjust the ionization and tautomerization state of the input molecules and predict probable forms at a specific pH, use the following command (replacing the pH value with your own choice:

```
$SCHRODINGER/epik -ph 7.0 -imae epik_input.mae
    -omae epik_prob_forms.mae
```

  The output structure file will contain predicted tautomers and ionized forms of the input molecules with a population greater than 0.01 at pH 7.0, with one or more structures corresponding to each input structure. Each output structure includes atom level properties for the pK_a value of that structure and four structure level properties that describe aspects of the overall likelihood for that structure existing. For more information on these properties, and how structural adjustment is performed in general, please see Chapter 5.

- Epik produces at least one output structure for each input structure, but occasionally the specific form in the input file may be deemed improbable and thus not appear in the output structure file. If you want to create alternate versions of the input structures, yet still retain the original forms, the following command can be used:

```
$SCHRODINGER/epik -retain_i -ph 7.0 -imae epik_input.mae
    -omae epik_retain_prob.mae
```

  The original ionization and tautomerization form for each input structure will be kept, and additional forms that Epik predicts with a population more probable than 0.01 are generated.

- To reduce the number of output structures produced, there are a number of options to choose from. For instance, the command:

```
$SCHRODINGER/epik -ph 7.0 -nt -p 0.1 -ms 2 -imae epik_input.mae
    -omae epik_2_ions.mae
```

  instructs Epik to skip tautomerization (`-nt`) and produce at most two (`-ms 2`) ionized forms for each input structure that have a probability greater than 0.1 (`-p`) at pH 7.0. If no ionized form has this high a probability, then the most probable ionized form is saved to the output file.

  To predict only the most probable form at pH 6.0, the following command can be used:

```
$SCHRODINGER/epik -ph 6.0 -ms 1 -imae epik_input.mae
    -omae epik_one_form_ph_6.mae
```

3.2.2 Distributing Epik jobs with para_epik

If you are running `epik` on many ligands, `para_epik` may be used to reduce the turnaround time by automatically distributing the calculation over multiple copies of `epik`. `para_epik` divides the collection of input structures into a specified number of sets (*njobs*). These sets are then used as input for several (*nprocs*) copies of `epik` until all of them have been processed.

To run the `para_epik` script from the command line, use the following syntax:

```
$SCHRODINGER/utilities/para_epik [options]
```

The options are described in Table 3.4. You can also use the same diagnostic options as for the `epik` command—see Table 3.3. Any other options or arguments you enter on the command line are passed to the `epik` command for each subjob. The Epik input structure files must be specified with `-imae` and the output structure file must be specified with `-omae`. The jobname for the overall `para_epik` run is derived from the output structure file name by removing everything after the last "." if present.

Table 3.4. Options for the `para_epik` *command, including job control options.*

Option	Description
`-DEBUG`	Show details of operation of the top-level script.
`-first` *firstlig*	First ligand to include. Default 1.
`-HELP | -h[elp]`	Print usage message and exit.
`-HOST` *host_list*	*host_list* is a list of one or more hosts on which to run the job. The list must be quoted if multiple hosts are specified: for example, `"hostname1:nprocs1 hostname2:nprocs2 ..."` Default: `localhost:1`
`-HOSTFILE` *host_file_name*	The name of the host file to use for this run.
`-INTERVAL` *sec*	The maximum time in seconds between updates of the *jobname*`.log` file.
`-j` *jobnum*	Subjob number to prepare. Default 0, meaning all subjobs.
`-JOBCTS` *maxctsjob*	Ensure that each subjob has no more than this many structures to process. Defaults to 10000.
`-last` *lastlig*	Last ligand to include. Default 0, meaning the last ligand the file.
`-LOCAL`	Do not use a temporary directory to store the files. Store files in the local directory.
`-NICE`	Run the job at reduced priority

Table 3.4. Options for the `para_epik` *command, including job control options. (Continued)*

Option	Description
`-NJOBS` *njobs*	Divide the overall job into *njobs* subjobs.
`-nx`	Just create input files for subjobs.
`-OUTPUT_ORG BY_SUBJOB`	Produce one output file for each subjob.
`-PROJ` *name*	Assign job to the Maestro project *name*.
`-TMPDIR` *directory*	Specify the directory used for temporary storage during a job.
`-v[er[sion]]`	Report the version number for `para_epik`
`-WAIT`	Do not return control to the shell until the job finishes.

3.3 Epik Limitations

Epik predicts pK_a values and generates ionization and tautomerization states for typical ligand-like organic molecules in a rapid and practical manner. Thus, some constraints are imposed on certain classes of systems, which Epik may not process as well as others.

- **Molecular representation:** Epik requires all-atom input structures to have the hydrogen atoms explicitly specified. While most of Epik's calculation facilities are coordinate-independent, some can depend on input geometry. Epik is expected to function acceptably with well-chosen 2D or 3D coordinates.
- **Molecule size limitation:** Epik's default molecular size setting does not adjust the structures of molecules larger than 150 atoms, which are atypically large for ligands. Molecules larger than 150 atoms may cause delays in processing times, due to the increase in the number of possible variations generated. The `-ma` option allows the adjustment of this setting to accommodate larger molecules.
- **Target molecule classes:** Epik is intended to function well and rapidly for typical organic ligand-like molecules. Many of the calculations performed in Epic use precalculated information that has been tabulated for fast lookup during the calculation. While the tabulations are extensive, they are not exhaustive, and some chemical functionalities may not have sufficient information for accurate treatment.
- **pK_a range:** Epik is intended to reliably predict the pK_a values for ionizable groups whose ionization state may change under the range of pH values most relevant for medicinal chemistry. For water this range is 4 to 10. For DMSO the range is less well defined but is approximately 4 to 30. To prevent protonation of atoms not normally regarded as basic due to simple, general, yet inappropriate matches for these atoms, more specific matches are used to assign very low, often negative pK_a values.

Epik Methodology

Empirical prediction of acid and base pK_a values for organic molecules has a long and largely successful history. The combination of two closely related linear free-energy approaches based upon the Hammett equation for aromatic molecules and the Taft equation for aliphatic molecules has been adopted for use in Epik to predict the pK_a values of organic acids and bases. The implementation used largely follows that described in *pK_a prediction for Organic Acids and Bases* (Perrin, D.D.; Dempsey, B.; and Serjeant, E.P.; Chapman and Hall, London (**1981**)). This approach is briefly and functionally described in the next section. Schrödinger's tautomerization methodology is described in Section 4.2. General information on Epik's standard parameter set is available in Section 4.3, while information on customizing the parametrization is available in Section 4.4.

4.1 pK_a Prediction

4.1.1 Overview of Hammett and Taft Prediction of pK_a Values

Hammett and Taft equations are intended to predict microscopic pK_a values. That is, given a molecular protonation state, what are the pK_a values for the first addition or removal of a proton from the various ionizable functional groups? The same pK_a prediction for a functional group applies to both the acidic and conjugate base forms. Both the Hammett and the Taft equations have the same general form:

$$pK_a = pK_a^0 + CF - \sum_i \rho_i \sum_j \sigma_{i,j} \quad (1)$$

Hammett and Taft parameters, typically pairs of pK_a^0 and ρ values, are specific for each type of ionizable group. pK_a^0 values describe the unperturbed pK_a value for the ionizable group and ρ values describe the sensitivity of the ionizable group to substituents attached to the ionizable group at particular locations. For a description of the selection of the pK_a^0 value, see Section 4.1.2.

Since substitutions can occur at different locations relative to the ionizable group, more than one ρ value may be needed to make a prediction. However, one of these, the one paired with the pK_a^0 value selected for the ionizable group, plays a primary role.

The perturbing influences of most structural features, most commonly substituents, are described by σ values. A description of the method of selection and estimation of the various σ values is found in Section 4.1.3. The remaining adjustments are described by a general correction factor, CF.

$$CF = -\log_{10}(n_{HR}/n_{HA}) + RA \quad (2)$$

The first term in Equation (2) accounts for the number of ways to remove equivalent H atoms from an acidic molecule (n_{HR}) versus the number of equivalent ways to add an H atom to the conjugate base (n_{HA}). The second term, RA, is an empirical ring adjustment term for aliphatic rings. For instance, a value for RA of roughly 0.2 is often used for ionizable amine atoms in a single aliphatic ring, such as morpholine.

Each pair of pK_a^0 and ρ values has an uncertainty associated with it that reflects the standard deviation versus experiment in the predictions given by that match. If insufficient data is available to determine an uncertainty, a default value of 2.0 pK_a units is used. Functional groups that lie outside Epik's parametrization coverage may be matched with unsuitable Hammett or Taft parameters, resulting in predictions whose accuracy lies well outside the uncertainties given.

4.1.2 Selection of pK_a^0 and Primary ρ Values

Association of pairs of Hammett or Taft pK_a^0 and ρ values with particular functional groups is controlled by the `acid_base` blocks in the parameter file for the solvent being used (see Appendix A). Each `acid_base` block contains a SMARTS pattern for the acidic form from which Epik automatically generates the conjugate base pattern. The SMARTS patterns are used to identify ionizable functional groups in the molecules. The first atom in the acid SMARTS pattern in the `acid_base` block is always the acidic proton and the second atom is referred to as the base atom. The base atom is the first atom in the automatically generated SMARTS pattern for the conjugate base. The `acid_base` block lists which atoms in the SMARTS pattern may have substituents attached.

If a functional group is matched by only one `acid_base` block, then the pK_a^0 and ρ from that pattern are used for it. Usually, ionizable functional groups are matched by multiple blocks for related functional groups. In such cases, the `acid_base` block that has the highest `priority` value is identified as the primary match. If two or more blocks have the same priority value, the matches are compared in a pairwise manner until one emerges as the most suitable for the current molecule. During this comparison, the rules given below are used sequentially to determine which match is more useful. For simplicity, the two matches are referred to as *a* and *b*.

1. Matches containing aromatic groups are selected over matches that do not.
2. If the potentially most influential substituent for match *a* is contained within the pattern for match *b*, match *b* is judged to be more suitable.

   Specifically: for matches *a* and *b*, find the atom (A) bonded to the substituent atom closest to the base atom that has the highest atomic number and lies outside of the SMARTS pattern for this match. If the A_a atom lies within the SMARTS pattern of match *b*, match *b* is more suitable unless atom A_b also lies in the SMARTS pattern for match *a*.
3. Othewise, the paths through the molecule to reach the first substituent for each match are compared in the following order:
   a. If one path is completely contained within another path the match with the longer path is selected.
   b. If all of the atoms in the path from the base atom to the closest substituent location are sp^2 and traverse at most one aromatic ring for match *a*, while the same is not true for match *b*, then match *a* is judged to be more suitable.
   c. Select the match with the path with the longer set of consecutive sp^2 atoms starting from the base atom.
   d. Select the match with fewer `hetero_aromatic` atoms.
   e. Keep the match that occurs earlier in the solvent parameter file.

The pK_a^0 value from the primary match is used for pK_a^0 in the Hammett or Taft equation. The ρ from the primary match is used in Equation (1) unless substitution locations are present that cannot be covered by the primary match. In such cases, ρ values from other matches may be used or if no suitable ones are available, an appropriate ρ value will be calculated on the fly where possible. The use of various ρ values is covered in Section 4.1.3.

4.1.3 ρ Values

Each `acid_base` block's ρ value is intended to describe the perturbations introduced by adding substituents to specific atoms, enumerated within the block using the `subs_atoms` designator. The ρ value from the primary match (see Section 4.1.2) is first used for substituents attached to atoms explicitly listed in the `subs_atoms` lists for the primary match or any atoms in aromatic ring systems that are at least partially included in the `acid_base` SMARTS pattern.

In addition, the primary ρ value is used for non-substituent corrections such as:

- heteroaromatic atoms (non-carbon aromatic atoms) that are one of:
  i. non-carbon atoms that match general aromatic types in the SMARTS patterns ([a]) for the `acid_base` and are listed as `hetero_aromatic` in the primary match `acid_base` block.
  ii. atoms in portions of aromatic rings that are not listed in the SMARTS pattern that are part of a larger aromatic ring system that is at least partially listed in the SMARTS pattern for the primary match `acid_base` (e.g. a fused ring system, but not another ring singly-bonded to the primary aromatic ring).
- aromatic topological corrections:

  Corrections for known topological patterns in polycyclic aromatic ring systems for specific functional groups (e.g., those listed in Table 7.1 of Perrin et al.).

If the primary match has an aromatic ring that is fused to additional aromatic rings that are not contained within the primary match, the ρ from the primary match is used for attachments to the extended aromatic ring system.

If the atoms in the primary match have substituents that have not been covered by the primary match directly or by extending aromatic ring systems, other `acid_base` matches for this same functional group are examined to see if they have suitable explicit substituent locations. If so, the ρ values for the other patterns are used in Equation (1) for these contributions.

If there are still substituents that do not have a ρ value associated with them then the `acid_base` patterns are re-examined to see if they have substitution locations within the primary match that lie between the first atom in the substituent and the base atom. If more than one eligible `acid_base` pattern is found, the one with a substitution location closest to the first substituent atom is selected. If such locations are identified their ρ values are used to describe the contributions for these substituents and transmission effects are taken into account (see Section 4.1.4).

If all of the above criteria fail, then a general methodology for selecting a ρ value is used. The path in general will not involve aromatic atoms so the formula given by Perrin for aliphatic ionizable groups is used:

$$\rho = 0.8 \times 2^{2-h} \tag{3}$$

where h is the number of atoms between the substituent and the base atom.

4.1.4 σ Values for Substituents

In addition to the σ values for heteroaromatic atoms and topological patterns for rings mentioned earlier, σ terms come from substituents. The perturbing influence of a substituent depends on whether it is connected to an aliphatic or aromatic portion of the `acid_base` pattern being used. Table 4.1 lists the σ types that may be given for a substituent. These σ types, along with a SMARTS pattern describing the chemical nature of the substituent group, are given in a substituent structure in the solvent database (see Appendix A).

Table 4.1. σ types for substituents

Type	Used for
`sigma_star`	most aliphatic attachment locations
`sigma_ortho`	on aromatic rings where the substituent is ortho to the ionizable location
`sigma_meta`	on aromatic rings where the substituent is meta to the ionizable location
`sigma_para`	on aromatic rings where the substituent is para to the ionizable location
`special_sigma_ortho`	similar to sigma_ortho but only for a particular ionizable group
`special_sigma_meta`	similar to sigma_meta but only for a particular ionizable group
`special_sigma_para`	similar to sigma_para but only for a particular ionizable group

We defer to the viewpoint expressed by Perrin that `special_sigma_para` constants can be used to describe how `sigma_para` values can vary depending on the ionizing group and do not support the `sigma_para`$^-$ formalism sometimes used in Hammett equations.

In general, parameterized `sigma_ortho` values are uncommon and often do depend on the ionizable group in question, so `special_sigma_ortho` group values are not unusual.

For some unsaturated aliphatic systems, `sigma_ortho`, `sigma_meta`, or `sigma_para` (or combinations thereof) are used if the underlying Taft parameterization was performed using them rather than σ^* exclusively. If so, the non-zero weights for these various σ types are indicated in the `acid_base` block using explicit `sigma_star_wt`, `sigma_ortho_wt`, `sigma_meta_wt` or `sigma_para_wt` identifiers (see Appendix A).

In addition to sometimes having their own σ types, non-carbon atoms in five-membered aromatic rings are conceptually replaced by two aromatic carbon atoms creating a virtual six-membered ring. Determination of the ortho, meta, or para designations for substituents and hetero atoms within the ring relative to the ionizable group is done in this virtual ring.

In the absence of explicit `sigma_star_wt`, `sigma_ortho_wt`, `sigma_meta_wt`, or `sigma_para_wt` designations for the `acid_base` match, Epik uses the methods presented by

Perrin et al. in section 7.2 to provide ways to generate σ values for aromatic ring systems if at least two types of aromatic σ values are given. For instance, in six-membered rings, if any two of the three σ values are provided, the other may be very roughly estimated (if needed) using those methods. For polyaromatic ring systems, this methodology relies on charges and geometries from template ring systems. Epik has a few of the most important ring systems encoded. The molecule's aromatic ring system is converted into a virtual one by expanding real five-membered rings to six-membered rings when hetero atoms are present. The largest template ring system that matches the virtual one and contains the attachment location for the ionizable group is used to provide ring charges and geometry.

For substituents that are one aromatic bond or two aromatic bonds away from the ionizable group, a pure σ_o or σ_m (or their special version) is used if available. Otherwise a mixture of σ types are used (if possible) to estimate the influence of the substituent depending on the distance between the ionizable group and the substituent locations in the virtual ring system as well as the charges in the template ring system.

Epik has an extensive set of σ types for many organic substituents. When examining a substituent group at a particular location there may be more than one type of substituent description in the solvent database that matches it. When this occurs, Epik chooses the substituent description with the most atoms in its SMARTS pattern. If two such substituent descriptions have the same number of atoms, the first one in the solvent database file is used.

In some cases, even the largest matching substituent description does not cover all of the substituent atoms in that branch of the molecule. For aliphatic substitution locations, Epik can correct for the influence of these additional atoms. The methodology used for these cases involves treating these additional atoms as substituents whose influence needs to be propagated back to the substitution location for the `acid_base` group being used. The σ used in Equation (1) is given by:

$$\sigma_{use} = \sigma^{*}\tau_{net} \tag{4}$$

σ^{*} is given for the additional substituent and τ_{net} is the net transmission coefficient of the atoms (k) between the substituent and the substitution location used in the `acid_base` match used for this branch of the molecule.

$$\tau_{net} = \prod_{k} \tau_{k} \tag{5}$$

where τ_k is the transmission coefficient listed in the solvent parameter file for atoms like atom k.

If τ_{net} drops below 0.0025 the contribution from the substituent groups is neglected.

This same approach for transmitting the influence of a substituent across intervening atoms is also used to generate effective σ values for auxiliary `acid_base` groups whose substituent locations lie within the primary match.

4.2 Tautomerization

Tautomers are an important class of isomers that can interconvert under physiological conditions. Tautomeric forms have different chemical properties and interact differently. For instance, one tautomeric form may interact with the active site of a protein more strongly than the other forms. Therefore, for some types of calculations, such as docking with Glide, considering the appropriate tautomeric forms of ligands can be important.

There does not seem to be a universal definition for tautomers. In Epik, tautomers are defined as isomers that meet the following conditions:

- In aqueous solution tautomers interconvert rapidly enough to be present as mixtures.
- One or more hydrogen atoms are bound to different atoms and the orders of one or more of the bonds between non-hydrogen atoms differs between tautomers.
- The non-hydrogen atom topology of the structure does not change during these interconversions.

Epik's definition of tautomers therefore excludes the aldose-hemiacetal ring opening and closing equilibrium in sugars that are sometimes regarded as tautomerizations. An example of the well-known keto-enol tautomerization is shown in Figure 4.1.

The tautomerization facility of Epik is intended to generate tautomers for input structures in a practical and flexible manner. It relies on a database of tautomeric templates to guide it in generating tautomers. Flexibility is provided in the command-line options and by permitting users to modify the database. For more information on this database, see Appendix B.

The tautomerization facility is not intended to generate all possible tautomers. The collection of groups of tautomers in the database is not exhaustive. As well, each set of tautomeric forms in the database is limited to those tautomers that are expected to have significant populations in aqueous solution, rather than being a comprehensive list. Tautomers in the database are assigned probabilities to assist in focusing on the most highly populated tautomeric forms.

The tautomerization of each structure can be divided into two stages: tautomer pattern matching and structure generation. These stages are described in the following sections.

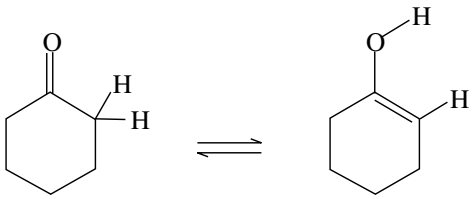

Figure 4.1. Keto-enol tautomerization.

4.2.1 Tautomer Pattern Matching

The current structure is examined for all matches of all tautomeric patterns in the tautomer database. All matches are cross-examined to see if the atoms are completely contained within another match. If so, the match containing the smaller number of atoms is eliminated. If they are the same size, then the match pattern found first is retained. This is the only case where the order of tautomers within a tautomer set or the order of the tautomer sets within the tautomer database affects processing.

4.2.2 Structure Generation

For each match, all tautomers in the tautomer set should be considered. Structures in which more than one match has been found have multiple locations in the structure that can be tautomerized, and all combinations of tautomers that are compatible should be considered. Double bonds that can shift to single bonds in any tautomer within a set can shift between E and Z forms, provided that other topological constraints do not prevent this shift.

The `tautomerizer` tries to generate all compatible combinations of all tautomers for each match. For each structure having a high enough probability, all combinations of 180 degree rotations about certain double bonds within each tautomeric pattern are generated. Double bonds are rotated if they involve C or N atoms in which both ends are not attached to two H atoms and in which the two atoms in the bond do not reside in the same ring. No adjustment is made to the probability for the tautomeric form for these double bond rotations. The probability for each structure resulting from this process is estimated as the product of the probabilities for all the tautomer templates used to generate it. The probabilities are normalized amongst all the structures generated for a given input structure.

Tautomers are generated in order of decreasing probability. By default, the maximum number of tautomers generated internally is 128. Of this collection of tautomers, up to eight tautomers are recorded by default in the output file for each input structure. This limit can be adjusted with the `-tn` option of the `epik` command. The most probable tautomer is always recorded. Except for this tautomer, only tautomers with a probability higher than 0.01 are recorded in the

output structure file. This threshold value can be adjusted using the `-tp` option of the `epik` command. If any of the `-retain` options is used to keep a tautomeric form of a molecule that would normally have a probability lower than the threshold probability value, the probability for this form is set to the threshold value.

Note: The probability for forms of a tautomerizable group is not adjusted for other functional groups inside the molecule, but outside the pattern itself. This means that these probabilities, and thus the overall probabilities for molecular forms (particularly when multiple tautomeric sites are being adjusted), are approximate and intended as a guide.

A chiral atom in a tautomer can switch its chirality if one of its tautomeric forms involves a double bond to this atom. Epik does not vary the chiralities of such atoms and does not add chirality properties for such atoms. Chiralities can be varied with the `stereoizer` utility in LigPrep—see Section 5.10 of the *LigPrep User Manual* for more information.

4.3 Standard Parameters

Epik comes with an extensive set of pK_a parameters built into the program. In addition, the parameter files `pKa_water_HT_data` and `pKa_DMSO_HT_data` in `$SCHRODINGER/mmshare-v`*version*`/data/mmpKa` may be used to update the parameter sets without needing to update the executable itself.

The tautomer database is also built into Epik. There is no standard location for tautomer definition updates. The tautomer database for DMSO is just a copy of that for water.

4.4 Custom Parameters

You can use custom parameter files to add to or override the standard parameters in Epik. The process is somewhat involved and the interaction between various data structures can be complex. This process may be changed in future releases.

Custom pK_a parameters and tautomer parameters are handled separately and differently. The format for pK_a parameter specification is described in Appendix A. The format for the tautomer database is described in Appendix B.

For pK_a parameters, information can be added to the files `pKa_water_HT_data` and `pKa_DMSO_HT_data` in `$SCHRODINGER/mmshare-v`*version*`/data/mmpKa`, or to a file specified using the command line option `-es`. If the latter method is used, it is advisable to create the file starting from a copy of the appropriate file from `$SCHRODINGER/mmshare-v`*version*`/data/mmpKa`, as specifying a file on the command line causes Epik to skip reading the file located in `mmshare`.

Including the `clear_standards` data item completely eliminates pre-existing pK_a parameters from the calculations. Using the `turn_off` data block overrides individual data structures. See Appendix A for more information on the use of these data items.

Custom tautomer parameters are provided to Epik through the `-ts` command line option. The default action of the `-ts` option adds the provided parameters to Epik's standard tautomer parameters. To clear out the standard parameters, use the `clear_standard` data item in the custom tautomer file.

Structural Adjustment in Epik

In addition to estimating the pK_a values for a given structure, Epik can adjust structures to attempt to generate a collection of structures consistent with the pH, while eliminating minimally contributing structural variations. The adjustment process involves both tautomerization and ionization.

In this chapter, we loosely use the phrase "population of a molecular structure" to mean a rough estimate for the fraction of a collection of related structures that would adopt this particular structure at equilibrium. We also use it at intermediate stages of adjusting structures for the equivalent quantity within some set of candidate structures.

Tautomerization is dependent on the ionization state of the molecule. As well, ionization depends on estimates for the pK_a of the functional group being considered for ionization, which in turn depends on the ionization states of other groups in the molecule, and of course, on the tautomeric state. Therefore, to obtain representative structures for a particular pH, these interdependencies must be taken into account. In Epik, structural adjustment is carried out iteratively. The first iteration starts with just the input structure, and subsequent iterations are applied to the collection of structures derived from the input structure in the preceding iterations. Each iteration consists of an attempt to tautomerize all structures in the current collection of molecules followed by an attempt to ionize all ionizable functional groups in each tautomer. Up to 5 iterations are attempted. This means, however, that at most 5 functional groups in the input molecule will have their ionization state adjusted. During each iteration, Epik tries to generate most tautomeric and ionization variations on the current collection of structures whose populations exceed a population threshold (*minpop*), up to a specified maximum number of structures (*ms*). If the number of structures exceeds *ms*, then the `ms` structures with the highest populations are produced. By default, *minpop* is 0.01 and is adjustable using Epik's `-p` option (or indirectly by using the `-pht` option). The default value for *ms* is 16, and is adjustable using the `-ms` option. Section 5.1 describes how the populations of the various tautomeric and ionic forms are estimated. Section 5.2 provides a more detailed description of the structural adjustment process.

5.1 Penalties and Populations for Ionization State and Tautomeric Forms

Tautomerization methodology and the estimation of tautomer populations are described in Chapter 7 of the *LigPrep User Manual*. pK_a values are estimated using the methods described in Chapter 3 of this manual. pK_a penalties for each structure are estimated using the methods described in Chapter 6 of the *LigPrep User Manual*, using Epik's predicted pK_a values. These penalties, which are expressed in kcal/mol, are recorded with each structure as shown in Table 5.1.

Table 5.1. pKa penalties for generated structure

Property Name	Description
r_epik_Ionization_Penalty	overall penalty for the structure
r_epik_Ionization_Penalty_Charging	penalty for having ionizable groups charged
r_epik_Ionization_Penalty_Neutral	penalty for having ionizable groups neutral
i_epik_Tot_Q	total charge of the structure

The `r_epik_Ionization_Penalty` property can be used to provide a rough estimate for the weight to assign to a particular ionization state, k, within the collection of the various ions generated from the same tautomeric form, t, using the following equation:

$$\mathrm{Wi}_k = e^{(-(\mathtt{r\_epik\_Ionization\_Penalty}_k)/k_B T)} \quad (1)$$

The calculation of populations for a specific combination of tautomeric and ionic forms of a molecule is a more complicated process since both tautomeric and ionic probabilities need to be accounted for at the same time. This process needs to be considered in the context of the interactive generation of a collection of structures, and is described in the following section.

5.2 Creating Structural Variations and Estimating Populations

The process for generating collections of tautomeric and ionic structural variations on the input structure is iterative, with each iteration involving a tautomeric stage and an ionization stage. In describing the procedure for adjusting the structures generated from a single input structure, it is helpful to use the following phrases:

- "current collection of structures" describes the structures generated in the previous stage of the process
- "new collection of structures" describes the set of structures being generated in the current stage of the process.

At the start of the first iteration, the current collection of structures is the input structure which is initially assigned a population of 1.0.

During the tautomerization stage of each iteration, each structure, *j*, in the current collection is tautomerized separately. The resulting tautomers are assigned weights, $Wt_{j,t}$, given by the following equation:

$$Wt_{j,t} = Pc_j p_t \tag{2}$$

where:

- Pc_j is the population of the structure *j* in the current collection
- p_t is the population of the tautomer divided by the population of the original tautomer

When the tautomers have been generated for all the current structures, the $Wt_{j,t}$ are normalized to give populations, $pt_{j,t}$ as shown in the following equation:

$$Pt_{j,t} = Wt_{j,t} \Big/ \left(\sum_{j,t} Wt_{j,t} \right) \tag{3}$$

All tautomeric forms with population weights lower than *minpop* are eliminated and the resulting structures are now considered the current collection of structures. If more than *ms* structures are present and *ms* 3, then only the *ms* most probable structures are kept. If *ms* < 3, three structures are kept.

All structures in the current collection of structures are now subjected to ionization. The pH range, *del_pH*, is given by:

$$del\_pH = -\log_{10}(minpop) \tag{4}$$

The following steps are carried out separately for each of the current structures, *j*:

1. Estimate the pK_a values for all ionizable groups
2. Accumulate structures into the new collection by examining each ionizable group in turn:
    i. If it is in a basic form and pK_a < pH + del_pH, the basic form is added to the new collection of structures.

ii. If it is in a basic form and pK_a > pH – del_pH, a proton is added to generate a new structure which is added to the new collection.

iii. If it is an acidic form and pK_a > pH – del_pH, the acid form is added to the new collection of structures

iv. If it is in an acidic form and pK_a < pH + del_pH, the proton is removed to generate a new structure which is added to the new collection.

If a new structure is the same as one already present in the new collection, it is not added a second time.

When all of the structures in the current collection of structures have been ionized, the weights for the new states are estimated:

$$\mathrm{Wn}_{j,k} = \mathrm{Pt}_j \mathrm{Wi}_k \tag{5}$$

which are normalized to give estimates for the populations:

$$\mathrm{Pc}_{j,k} = \mathrm{Wn}_{j,k} \Big/ \left(\sum_{j,k} \mathrm{Wn}_{j,k} \right) \tag{6}$$

Those with a population lower than the *minpop* threshold are eliminated. If there are more than *ms* structures, only the *ms* structures with the highest population are retained, with a minimum of 3 structures. The resulting set of structures is now considered the current collection of structures. The populations of individual structures in this collection, *j*, are now referred to as Pc_j.

For each iteration, the process of tautomerizing and ionizing is repeated until either the collection of structures does not change from one iteration to the next, or the maximum number of iterations (5) is reached. At the end of this process, using the population for each structure, Pc_j, the overall state penalty for each structure is calculated using:

$$\texttt{r\_epik\_State\_Penalty}_j = -k_B T \ln(\mathrm{Pc}_j) \tag{7}$$

Please note that the process used to determine `r_epik_State_Penalty` involves a number of significant approximations and thus, this quantity should only be used as a rough guide for gauging the importance of the output structures.

Getting Help

Schrödinger software is distributed with documentation in PDF format. If the documentation is not installed in `$SCHRODINGER/docs` on a computer that you have access to, you should install it or ask your system administrator to install it.

For help installing and setting up licenses for Schrödinger software and installing documentation, see the *Installation Guide*. For information on running jobs, see the *Job Control Guide*

Maestro has automatic, context-sensitive help (Auto-Help and Balloon Help, or tooltips), and an online help system. To get help, follow the steps below.

- Check the Auto-Help text box, which is located at the foot of the main window. If help is available for the task you are performing, it is automatically displayed there. Auto-Help contains a single line of information. For more detailed information, use the online help.
- If you want information about a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- For information about a panel or the tab that is displayed in a panel, click the Help button in the panel. The Help panel is opened and a relevant help topic is displayed.
- For other information in the online help, open the Help panel and locate the topic by searching or by category. You can open the Help panel by choosing Help from the Help menu on the main menu bar or by pressing CTRL+H.

  To view a list of all available Epik–related help topics, choose Epik from the Categories menu of the Categories tab. Double-click a topic title to view the topic.

If you do not find the information you need in the Maestro help system, check the following sources:

- *Maestro User Manual*, for detailed information on using Maestro
- *Maestro Command Reference Manual*, for information on Maestro commands
- Frequently Asked Questions pages, at https://www.schrodinger.com/Epik_FAQ.html

The manuals are also available in PDF format from the Schrödinger Support Center. Information on additions and corrections to the manuals is available from this web page.

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail: help@schrodinger.com
USPS: Schrödinger, 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information, most of which can be obtained by entering `$SCHRODINGER/machid` at a command prompt:

- All relevant user input and machine output
- Epik purchaser (company, research institution, or individual)
- Primary Epik user
- Computer platform type
- Operating system with version number
- Epik version number
- Maestro version number
- mmshare version number

pK_a Data File Format

This appendix describes the encoding of pK_a information in a solvent parameter file for Epik. The default parameter files for water and DMSO are not user-accessible. You may use your own parameter files to add to, or replace, the default parameter sets for these solvents or to describe new solvents.

A.1 Basic Elements of the Parameter File

The solvent parameter file has three basic elements:

- comment line
- information lines
- data blocks

Any line beginning with a # symbol is regarded as a comment. Information is provided by lines containing *information_name*: *information_value* pairings.

```
name: Hammett_Taft_water
short_name: H20
```

Only one *information_name*: *information_value* pairing can appear on any line.

Data blocks associate information lines into distinct groups. For instance, the following data block specifies the information needed to describe the Taft relation for a tertiary alcohol:

```
acid_base{
        Sch_ID: 13
        name: Alcohols, tertiary
        pattern: [#1]-[OX2]-[CX4HO]
        identifier: alcohol_tertiary
        sub_atoms: 3
        pKa0: 15.9
        rho: 1.42
}
```

while this block contains Hammett information for protonating aniline:

```
acid_base{
        Sch_ID: 1029
        name: Anilinium ions
        pattern: [#1]-[NX4+](-[#1])(-[#1])-cl[a][a][a][a][a]1
        identifier: aniliniumIon_phenyl
        sub_atoms: 6 7 8 9 10
        hetero_aromatic: 6 7 8 9 10
        pKa0: 4.58
        rho: 2.88
}
```

Blocks must begin with `block_type_name{` and end with a corresponding `}`. Only one occurrence of any *information_name* can appear within a given block.

Information lines that lie outside a data block apply to the whole solvent description, while information lines that lie inside blocks apply only to instances where the block is used. Aside from block-level grouping of information, data can be provided in any order in the solvent parameter file.

Table A.1. Data types that occur outside a block

Information_name[a]	Type	Description
`name`	string	Full name for entire set of pK_a data
`short_name`	string	Shorter name for entire set of pK_a data
`temperature`	real	Temperature for the parametrization in degrees C.
`clear_standard`	none	Clears out pK_a parameter information present before reading this file.

a. All of these descriptors are required in the parameterization file.

There are a number of types of top-level data blocks:

- `acid_base` — Describes a distinct parameterization for an acid/conjugate base pair.
- `transmission` — definition of the atom in a transmission group, and how strongly it propagates the perturbations .
- `substituent` — Description of substituent fragments and their perturbing effects.
- `ring_adjustment` — Describes how to adjust pK_a values for base atoms (the second atom in acid_base patterns) in aliphatic rings.
- `hetero_atom_group` — Describes hetero atom based groups in aromatic rings for which automatic adjustment of the pK_a should be attempted.

- `hydrogen_penalty` — Describes which hydrogen σ types to use when removing hydrogens from non-explicit substitution locations. Only one `hydrogen_penalty` should be in the file.
- `turn_off` — Lists previously defined data blocks to disable.

The following information_name values describe special σ types, and can only appear within substituent blocks:

- `special_sigma_ortho`
- `special_sigma_meta`
- `special_sigma_para`

Table A.2 provides a description of the types of data blocks.

Table A.2. Data block component definitions

Information_name	Required	Type	Purpose
`acid_base`			
`name`	Yes	String name	Name for ionizable functional group.
`Sch_ID`	No	Integer	For identification.
`pattern`	Yes	String	SMARTS pattern for acidic form.
`identifier`	No	String	Used to systematically describe the nature of the group.
`subs_atoms`	Yes	Integers	A list of atom numbers in the SMARTS pattern for which the ρ value given for the attachment of substituents is appropriate.
`hetero_aromatic`	No	Integers	A list of atom numbers within the SMARTS pattern for which automatic corrections may be applied for hetero_atoms defined in the hetero_atom_group blocks may be used.
`pKa0`	Yes	Real	Specifies the Hammett or Taft pK_a^0 value.
`rho`	Yes	Real	Specifies the Hammett or Taft ρ value.
`pKa_unc`	No	Real	The uncertainty in the pK_a values estimated for this group. Default is 2.0.

Table A.2. Data block component definitions

Information_name	Required	Type	Purpose
use_hydrogen_sigma	No	Integer	Indicates whether hydrogen σ values are used. Default is 0 (no); yes is 1.
priority	No	Real	Assigns a priority to the group. Default is 0. Larger values denote a high precedence.
sigma_ortho_wt	No	Real	Normally, which σ value to use is determined by the nature of the molecule. However, some parametrizations for particular aliphatic functional groups explicitly state that certain non-default σ values must be used.
sigma_meta_wt	No	Real	
sigma_para_wt	No	Real	
sigma_star_wt	No	Real	
transmission			
name	Yes	String	Name for the transmission group.
Sch_ID	No	Integer	For identification
pattern	Yes	String	SMARTS pattern for the transmission group. Multi-atom transmission groups are not supported.
tcoef	Yes	Real	Transmission coefficient. τ typically 0.4 (range = 0.3 to 0.7)
substituent			
name	Yes	String	Name for substituent.
Sch_ID	No	Integer	For identification.
pattern	Yes	String	SMARTS pattern for substituent group.
sigma_ortho	At least one σ must be given	Real	Influence of substituent (> 0 acid strengthening)
sigma_meta		Real	
sigma_para		Real	
sigma_star		Real	
special_sigma_ortho	No	Structure	See special_sigma structure. Used to give special σ values for combinations of certain substituents of aromatics with certain acid/base groups.
special_sigma_meta	No	Structure	
special_sigma_para	No	Structure	

Table A.2. Data block component definitions

Information_name	Required	Type	Purpose
special_sigma_ortho special_sigma_meta special_sigma_para			
name	Yes	String	Name of acid for special σ value.
Sch_ID	No	Integer	For identification.
pattern	Yes	String	Pattern for acidic version of special acid/base substituent combination.
sigma	Yes	Real	σ value to use.
ring_adjustment			
name	Yes	String	Name of ring adjustment.
Sch_ID	No	Integer	For identification.
pattern	Yes	String	SMARTS pattern, usually just one atom (e.g., N).
pKa_shift1	Yes	Real	Shift if in 1 ring.
pKa_shift2	No	Real	Shift if in 2 rings.
hetero_atom_group			
name	Yes	String	Name of hetero atom group.
Sch_ID	No	Integer	For identification.
pattern	Yes	String	SMARTS pattern to match. First atom should be in the aromatic ring.
sigma_ortho	One of these three must be included	Real	σ adjustments for different ring locations.
sigma_meta		Real	
sigma_para		Real	
ring_size	Yes	Integer	Apply only to rings of this size.
hydrogen_penalty			
sigma_ortho	One of these four must be included	Real	σ value to use.
sigma_meta		Real	
sigma_para		Real	
sigma_star		Real	

Table A.2. Data block component definitions

Information_name	Required	Type	Purpose
`turn_off`			
`acid_base`	No	Integer	`Sch_ID` for structure to turn off.
`transmission`	No	Integer	`Sch_ID` for structure to turn off.
`substituent`	No	Integer	`Sch_ID` for structure to turn off.
`ring_adjustment`	No	Integer	`Sch_ID` for structure to turn off.
`hetero_atom_group`	No	Integer	`Sch_ID` for structure to turn off.

A.2 Notes about SMARTS patterns for acid_base and special_sigma blocks

The SMARTS patterns in these blocks have some restrictions placed on them. Patterns must show the acidic form. The conjugate base form is generated automatically from the acidic SMARTS pattern. The first atom must be the acidic hydrogen and be represented as: [#1]. The second atom in the pattern must be the base atom (the atom to which the acidic hydrogen is bonded). This atom may have the X, H, v and +/- (charge) qualifiers; if more than one of them is present, they must appear in this order (e.g., NX3H2v4+).

A.3 Notes about acid_base groups

The parameter file must contain at least one `acid_base` group. Multiple groups may match for a particular ionizable location. One is selected based upon its priority, size, proximity of substitution locations to base atom, and extent of unsaturation between substitution locations and the base atom. The pK_a^0 from this pattern is used. ρ values from multiple patterns may be used. See Chapter 4 for more information.

Tautomer Database Format

In Epik 1.0 the default tautomer database is not accessible to users. However, you can provide your own file to either completely override or add patterns to the default tautomer collection.

At the top level of the tautomer database file the following four items can be present: `name`, `clear_standard`, `group_def`, and `tautomer_set`. These items are described in the following sections. Lines beginning with a `#` are comment lines and are ignored when interpreting the contents of the tautomer database file. Blank lines are also ignored.

name Data Item

`name` specifies the name of the solvent. For example:

```
name: water
```

`water` and `DMSO` are standard names for which the tautomerizer already has information. Currently, the DMSO tautomer information is just a duplicate of that for water.

clear_standard Directive

By default, information in a custom tautomer database file is added to any existing information already available for the solvent specified. Including `clear_standard:` in a tautomer database file clears any values for this solvent accumulated before the current file was read.

group_def Data Structure

The tautomerization facility does not support recursive SMARTS. However, a mechanism that supports some of the functionality of recursive SMARTS is provided by the `group_def` data structure. This data structure permits you to define variables that correspond to SMARTS patterns. The variables may be reused in groups and `tautomer_sets` that appear later in the tautomer database file.

Each group contains two items:

`name:` an arbitrary name for the group which is used to reference the group.

`pattern:` The SMARTS pattern for the group. This pattern may refer to previously defined groups using $*groupname*.

Below are some examples of `group_def` data structures:

```
group_def{
        name: Halogens
        pattern: [F,Cl,Br,I]
}

group_def{
        name: Amides
        pattern: [CX3](=[OX1])-[NX3]
}

group_def{
        name: Carbonyls
        pattern: [CX3](=[OX1])
}

group_def{
        name: Carbonyls_only
        pattern: [$Carbonyls;!$Amides]
}
```

tautomer_set Data Structure

`tautomer_set` data structures define sets of interconvertible tautomers. There are more than 150 tautomer sets available by default for water.

Some examples of `tautomer_set` data structures are given below, and the syntax for the data structures is described following these examples.

Note: The entry for `pattern:` values must be a single line. In the examples below, some of the `pattern:` text wraps to the next line due to formatting constraints within this manual. When creating tautomer data structure files in a text editor, ensure that text-wrapping is turned off, or that margins are set wide enough to accommodate single-line entry for this value.

```
tautomer_set{
    name: single-sided_ket-enol
# From: Handbook of organic chemistry

    tautomer{
      name: enol
      pattern: [CX3](-[#1,$Sub_aC])(-[#1,$Sub_aC])=[CX3](-
[#1,$Sub_carbonyl_C])-[OX2]-[#1]
      probability: 0.00005
    }

    tautomer{
      name: ket
      pattern: [CX4](-[#1])(-[#1,$Sub_aC])(-[#1,$Sub_aC])-[CX3](-
[#1,$Sub_carbonyl_C])=[OX1]
      probability: 0.99995
    }
}

tautomer_set{
    name: double-sided_ket-enol
# From: Handbook of organic chemistry

    tautomer{
      name: 1enol
      pattern: [CX3](-[#1,$Sub_aC])(-[#1,$Sub_aC])=[CX3](-[CX4](-
[#1])(-[#1,$Sub_aC])(-[#1,$Sub_aC]))-[OX2]-[#1]
      probability: 0.00000001
    }

    tautomer{
      name: ket
      pattern: [CX4](-[#1,$Sub_aC])(-[#1,$Sub_aC])(-[#1])-[CX3](-
[CX4](-[#1])(-[#1,$Sub_aC])(-[#1,$Sub_aC]))=[OX1]
      probability: .99999998
    }

    tautomer{
      name: 2enol
      pattern: [CX4](-[#1,$Sub_aC])(-[#1,$Sub_aC])(-[#1])-
[CX3](=[CX3](-[#1,$Sub_aC])(-[#1,$Sub_aC]))-[OX2]-[#1]
      probability: 0.00000001
    }
```

```
}
tautomer_set{
    name: imidazole

    tautomer{
      name: form1
      pattern: c1(~[#1,$Sub_c])n(-[#1,$Sub_n])-c(-
[#1,$Sub_c])=[nX2]c1(~[#1,$Sub_c])
      probability: 0.50
    }

    tautomer{
      name: form2
      pattern: c1(~[#1,$Sub_c])[nX2]=c(-[#1,$Sub_c])-n(-
[#1,$Sub_n])c1(~[#1,$Sub_c])
      probability: 0.50
    }
}
```

Each tautomer set contains a `name:` designator and a number of tautomer structures. The `name:` designator is followed by a space and a contiguous non-blank label to identify the class of tautomers described by the set. The label provided does not affect processing. In the examples below, there are three tautomeric sets: `single-sided_enol-ket`, `double-sided_enol-ket`, and `imidazole`.

The `tautomer` structure describes the properties of one tautomeric form. There are three designators that may be used within a `tautomer` structure: `name:`, `probability:`, and `pattern:`.

The `name:` designator provides a label for the tautomer but does not otherwise affect processing.

The `probability:` designator is used to assign a probability or fractional population of this tautomer within this tautomeric set. In many cases, reliable information on the probability of various tautomeric forms is not available and the values entered in the database are simply educated guesses.

The `pattern:` designator is followed by a contiguous SMARTS-like pattern. A difference between this pattern and a normal SMARTS pattern is that explicit single "–" and double "=" bond designators are used to make the corresponding Lewis structures clear. In addition, these patterns may include references to previously defined groups via the $*group_name* mechanism. Information on SMARTS patterns is provided on the web page: http://www.daylight.com/learn[1]. The SMARTS-like pattern is used to detect the corresponding groups of molecules in the input structures and to permit the tautomerization facility to under-

stand how the bonding patterns (Lewis structures) differ between tautomers so that they may be interconverted. For heavy atoms that are expected to carry a formal charge it is advisable to include the charge in the SMARTS pattern. To ensure that the SMARTS patterns are properly interpreted by Epik, the following restrictions must be applied:

- The SMARTS patterns for all tautomers within a tautomer set include the same list of non-hydrogen atoms in the same order.
- All SMARTS patterns must explicitly designate the hydrogens that shift positions in any tautomer within a tautomer set with a `-[#1]` pattern.
- All SMARTS patterns within a tautomer set must contain the same number of explicitly designated mobile hydrogen atoms.
- In both non-aromatic and aromatic portions of the SMARTS pattern, bond orders that change between single and double in any tautomer must be explicitly specified in the SMARTS patterns for all tautomers in a tautomer set.
- In portions of molecules that must be represented by aromatic atom types (e.g., `c` and `n`), only changes in the bond orders of bonds involving `n` atoms in the corresponding Lewis structures are supported. If such a bond changes order in any tautomer in a tautomer set, it must be represented as '`:`' in all the tautomers. See the guanosine tautomer set in the example above.
- Recursive SMARTS patterns are not supported.
- SMARTS patterns within the same tautomer set must all specify the same overall formal charge.

The database provided with this release contains templates for keto-enol tautomers and their sulfur analogues, imine-enamine tautomers, histidine-like tautomers, tautomers of DNA and RNA bases, and a large number of common heteroaromatic rings containing C, S, O, and N.

1. Please see the notice regarding third party programs and third party Web sites on the copyright page at the front of this manual.

Index

A

aliphatic substitution corrections 44
aromatic rings
  five-membered 43
  polycyclic 42, 44
  with heteroatoms 42
aromatic topological corrections 42
atoms, selecting 23–25
Auto-Help 30, 53

B

Balloon Help 30, 53
base atom, definition 40
Build panel 21
building structures 20–23
button menu 9

C

chiralities, change during tautomerization 47
Command Script Editor panel 26
command scripts—*see* scripts
conformations, double bond 46
conventions, document vii
correction factor 40
corrections
  aromatic topological 42
  non-substituent 42
current working directory 6

D

directory
  current working 6, 27
  output 27
  temporary 34
double-bond conformations 46

E

entries, Project Table 13
  including, excluding, and fixing 17
  selecting 17
  sorting 15
environment variables
  `DISPLAY` 6
  `SCHRODINGER` 5–6
ePlayer 15, 16
excluded entries 17

F

file I/O directory 27
filters, project entry 17
five-membered aromatic rings 43
fixed entries 18
fragments, building structures from 20
full screen mode 8, 13
function key macros—*see* scripts

G

grow bond 21

H

Hammett equations 2
  general form 39
  parameter derivations 2
Help panel 30, 53
hosts file 34

I

included entries 17
input structures, requirements on 38
ionization state
  keeping original 32, 33
  population of 52

J

jobs, running in Maestro 28–29

L

log file, saving Maestro 30

M

macros—*see* scripts
Maestro
  main window 6, 7
  menus 8
  quitting 30
  running jobs from 28–29
  scratch projects 13
  starting 6
  undoing operations 28

main window 7
menu button 9
minimum probability 33
Monitor panel 29
mouse functions 5
  Project Table panel 18–19
  Workspace 12

N

non-substituent corrections 42

O

online help 30
options
  `para_ligprep` 37

P

`para_ligprep`
  options 37
pH range 32, 33
pH target value 33
pK_a values
  display of 31
  storage of 31
  unperturbed 39
population of a structure, definition 49
pre-existing parameters, overriding 48, 61
Preferences panel 27, 28
product installation 53
project entries, *see* entries, Project Table
Project Facility, introduction 13
Project Table panel 15
  menus 16
  mouse functions 18–19
  shortcut keys 19
projects 13
Python scripts—*see* scripts

Q

quitting Maestro 30

R

recursive SMARTS 61
ring adjustment term 40

S

Schrödinger contact information 54
`schrodinger.hosts` file 34
scratch entries 14
scratch projects 13
scripts
  function key macros 27
  macros 27
  Maestro command 26
  Python 25
selecting objects in the Workspace 9, 23
sensitivity to substituents 39
shortcut keys
  main window 13
  Project Table panel 19
SMARTS patterns
  recursive 61
  restrictions in tautomer database 65
structures
  adjustment process 2
  building 20–23
  displaying in sequence 15
  maximum number generated 33
  maximum number of atoms 33
  minimum number kept 51
  requirements on 38
  selection of 3
  subset selection utilities 4
substituents
  choice of description 44
  data file description of 56
  path to 41
  sensitivity to 39

T

Taft equations 2
  general form 39
  parameter derivations 2
tautomerization, limitations on 45
tautomers
  definition 45
  generating 32
  keeping original 32, 33
  maximum number 33
  minimum probability 33
  pattern matching 45

screening by probability 46
types recognized 65
weight assignment 51
technical support 30
toolbar
Build panel..... 22–23
main window 9–12
Project Table panel 15–16
transmission coefficient 44, 58
transmission group..... 56

U

uncertainty values
determination of..... 40
storage of 31
undoing Maestro operations..... 28
utilities, additional 3

W

Workspace
description 6
full screen mode 8, 13
including, excluding, and fixing entries 17
mouse functions..... 12
scratch entries 14

120 West 45th Street
32nd Floor
New York, NY 10036

101 SW Main Street
Suite 1300
Portland, OR 97204

3655 Nobel Drive
Suite 430
San Diego, CA 92122

Dynamostraße 13
68165 Mannheim
Germany

QuatroHouse, Frimley Road
Camberley GU16 7ER
United Kingdom

SCHRÖDINGER.